# HANDBOOK

# ECONOMIC TENDENCY SURVEYS

***Draft – May 2014***

# PREFACE

# ACKNOWLEDGEMENTS

The *Handbook* was prepared by a working group led by the Italian National Institute of Statistics (ISTAT) and comprising the following contributors: ISTAT, the Organization for Economic Cooperation and Development (OECD), the Swiss Economic Institute (KOF-ETH Zürich), the Philippine Statistics Authority (PSA former NSCB), Statistics Netherlands (Central Bureau for Statistics-CBS).

The European Commission (EC) and the French National Institute of Statistics and Economic Studies (INSEE) have also reviewed draft chapters of the handbook.

Thanks to the following experts: Mauro Politi, Roberto Gismondi, Luciana Crosilla, Solange Leproux, Emilia Matera, Patrizia Margani, Paola Bellincampi, Bianca Maria Martelli (ISTAT); Marco Malgarini (Italian Agency for the Evaluation of Universities and Research Institutes - ANVUR); Sidney Vergouw, Michel van Kooten, Guido van den Heuvel. Roberto Wekker and Leo Enthoven (CBS); Klaus Abberger and Richard Etter (KOF-ETH); Cynthia Regalado (PSA); Michela Gamba, Pierre-Alain Pionnier and Gyorgy Gyomai (OECD).

Particular mention may be made of the contribution of Aloisio Campelo Junior and Viviane Seda Bittencourt (Instituto Brasileiro de Economia, Fundação Getulio Vargas, IBRE-FGV) on how Business and Consumer Tendency Surveys are conducted in Brazil.

We are also grateful for their useful comments and revisions to Christian Gayer, Roberta Friz, Andreas Reuter, Alessandro Girardi, (EC - DG ECFIN) and Cedric Audenis (INSEE).

Special thanks to Mauro Politi, Luciana Crosilla and Jolanda Giaconi (ISTAT) for their invaluable support to the coordination and finalization of the Handbook and to Gulab Singh, Ivo Havinga and Herman Smith (UNSD) for their work of support.

# CONTENTS

3. (*continued*)



4. **Survey frame and sample design**



**5. Estimation procedures and accuracy**

8. (*continued*)

# 10. Use of tendency survey results



*Figures*

# LIST OF ABBREVIATIONS AND ACRONYMS

| | |
|---|---|
| ARIMA | AutoRegressive Integrated Moving Average |
| BCI | Business Climate Indicator |
| BCS | Business and Consumer Surveys |
| BM | Bridge Models |
| BTS | Business Tendency Survey |
| CALMAR | CALages sur MARges (CALibration on MARgines) |
| CAPI | Computer Assisted Personal Interviewing |
| CASAQ | Computer Assisted Self-Administrated Questionnaire |
| CASI | Computer Assisted Self- Interviewing |
| CATI | Computer Assisted Telephone Interviewing |
| CCI | Consumer Confidence Indicator |
| CEPR | Centre for Economic Policy Research |
| CIRET | Centre for International Research on Economic Tendency Surveys |
| CLI | Composite Leading Indicator |
| CTS | Consumer Tendency Survey |
| DCM | Data Collection Mode |
| DG ECFIN | Directorate-General for Economic and Financial Affairs of the European Commission |
| EC | European Commission |
| ECB | European Central Bank |
| EPSEM | Equal Probability of Selection Method |
| ESI | Economic Sentiment Indicator |
| ETS | Economic tendency Survey |
| EU | European Union |
| FM | Factor Models |

| | |
|---|---|
| GDP | Gross Domestic Product |
| HT | Horvitz-Thompson estimator |
| ICI | Industrial Confidence Indicator |
| IESI | Italian ISTAT Economic Sentiment Indicator |
| IFO | Information and Forschung (Research). German Institute for Economic Research |
| INSEE | French National Institute of Statistics and Economic Studies |
| IPI | Italian index of industrial production |
| ISIC | International Standard Industrial Classification of All Economic Activities |
| ISTAT | Italian National Institute of Statistics |
| IT | Information Technology |
| KAU | Kind-of-Activity Unit |
| LAU | Local Administrative Units |
| LOCF | Last observation carried forward |
| MSE | Mean Squared Error |
| MSM | Markov Switching Models |
| NACE | Classification of economic activities in the European Community |
| NBER | National Bureau of Economic Research |
| NUTS | Nomenclature of Territorial Units for Statistics |
| OECD | Organisation for Economic Co-operation and Development |
| PAM | Partition Around Medoids |
| PDF | Probability Density Function |
| PIH | Permanent Income Hypothesis |
| PMI | Purchasing Managers' Index |
| PPS | Probability Proportional to Size |
| RCI | Retail trade Confidence Indicator |
| RDD | Random Digit Dialing |

| | |
|---|---|
| RGENESEES | R Evolved Generalised Software for Sampling Estimates and Errors in Surveys |
| RGH | Rational Expectations Hypothesis |
| SCI | Services Confidence Indicator |
| SEATS | Signal Extraction in Arima Time Series |
| SN | Statistics Netherlands |
| SRS | Simple Random Sampling |
| STAMP | Structural Time Series Analyser, Modeller and Predictor |
| STM | Structural Time Series Models |
| TRAMO | Time series Regression with Arima noise, Missing values and Outliers |
| UK | United Kingdom |
| UN | United Nations |
| US | United States |

# CHAPTER 1

# Introduction

## A. Main goals of the handbook

1.1 Economic Tendency Surveys (ETS) are well-established tools for the assessment and analysis of economic development and fluctuations in the business cycle. They have proven successful in many countries and under different economic and social conditions. ETS ask company managers and consumers about the current situation of their business/households, their opinion on the whole economy and about their plans and expectations for the near future. Generally, respondents are not asked for precise (quantitative) information on (for example) levels of output, sales, investment, employment and household finance. Instead they are asked to give qualitative information.

1.2 The first ETS originated in the 1920's, but underwent a real adoption and evolution since the 1950's. In Europe, the European Commission (EC), through the Directorate-General for Economic and Financial Affairs (DG ECFIN) launched a harmonization programme in 1961 and since 1996 tendency surveys are conducted in five areas (manufacturing, retail, construction, services and consumers)[1], conforming to a harmonized methodology in all 28 member states of the European Union (EU) and candidate countries. The Organisation for Economic Co-operation and Development (OECD) has collected some best practices and compiled a handbook in 2003. The Centre for International Research on ETS (CIRET) is a forum for compilers and users of ETS.

1.3 Survey results are used by various groups of customers. Among them are businesses, analysts, national and commercial banks, associations, policy makers and the media. They all need timely and reliable data for their analyses.

1.4 ETS results are seen as valuable in countries with a weak or developing system of quantitative official statistics but also in countries with well-established national accounts and other quantitative official statistics. In the former case ETS results are often one of a few or even the only source of high frequency economic data. However, even in countries with high quality national accounts ETS results are valuable additional information to other official data. Therefore even in large industrialized economies indicators like "Consumer Sentiment" or "Business Climate" are

[1] In the meantime also the banking sector is covered by the Programme.

closely watched by economic observers. Also the OECD system of early indicators relies in many countries on ETS data.

1.5 International seminars, like the United Nations (UN) seminars on rapid indicators and Eurostat Colloquium on business cycle analysis, for example, have recognised the usefulness of confidence indicators in providing an early warning of changes in economic activity. ETS data can function as stand-alone indicators, as tools to improve the estimation of other economic indicators, and as components of coherent indicator sets and composite indicators. They are thus an important part of the statistical system. However, more international harmonisation and improving the understanding of how confidence indicators and economic indicators can be brought together are desirable. UN has committed to establish a handbook with best practices and urges countries to conduct ETS, if not doing so already.

1.6 This distinguishes this handbook from various existing ones. The EC, within its harmonized survey programme, developed and described in 2007 (latest version March 2014) common standards and a set of harmonized questions in the user guide "A User Manual to the Joint Harmonized EU Programme of Business and Consumer Surveys". The OECD handbook "Business Tendency Surveys: A Handbook" (2003), is based on the experience gained by OECD staff in working with a wide range of countries on the implementation of harmonized Business Tendency Surveys (BTS) and the use of the results for short-term economic analysis. In this handbook, the scope is broader: it includes Consumer Tendency Surveys (CTS) as well and introduces also more flexibility for country specific surveys. However, for international comparability, each survey must apply some basic rules and questions.

1.7 This handbook describes some standards regarding the conduct of ETS, which are applicable in a great variety of economies. It provides all ingredients for institutes who intend to compile ETS data to set-up these surveys. Furthermore, the handbook offers to institutes which are conducting ETS some guidelines in improving their surveys and engaging in constructing derived indicators. On the other hand users of ETS can see what the guiding principles and the methods used behind the surveys are.

1.8 Why using common practices? A common understanding of best practices helps to ensure the quality and reliability of survey results. And a common set of questions and methods helps to simplify international comparisons. Thus trust regarding the instrument of ETS will be increased with domestic economic observers. Furthermore, the work of the international community is

disburdened, because they can easier interpret the survey outcomes. So standardization will increase the value of ETS results even further.

1.9 However, despite the importance of common practices, one approach doesn't fit completely for all circumstances. Regarding ETS this means that there could be particular sectors or branches important for the development of the domestic economy, which are not covered by the majority of countries. Also specific questions could be a useful complementation to internationalized standard questions. This might be justified by specific market conditions or by institutional, legal or infrastructure peculiarities. Regarding the survey process, adaptions are needed in the sampling procedure, the survey mode, the survey compilation. However, these adaptions, which are often needed to get results of higher quality, should be well-founded. This handbook presents principles guiding the conduct of ETS.

1.10 The audience or target users of this handbook are compilers of ETS, institutes who consider setting up a framework for conducting ETS (whenever conducting ETS or use ETS in their analysis) and compilers of composite indicators considering to include ETS. On the other hand, policymakers and anybody related to monitoring economic conditions and business cycles can benefit from the insights in the compiling of ETS and related products in this handbook.

## B. Qualitative data: characteristics and place in the system of economic statistics

1.11 ETS obtain *qualitative data* on a wide range of topics. Entrepreneurs and consumers are asked for past developments, assessments on current developments and expectations for the near future. For each question, the respondent has to choose between a limited number of pre-given answers, describing the past and current developments in qualitative statements. This practice is rather different from that of official statistics, which ask for a numeric evaluation of the variable concerned. Thus all economic statistics are considered *quantitative data*.

1.12 ETS have some unique characteristics. The most important feature of ETS is its timeliness. The simple and intuitive questionnaire makes it very easy for the respondent to complete it quickly and within or relatively shortly after the end of the reference period. There is no need to extract exact figures from the administration, just an indicative statement on developments and expectations is sufficient. Because of this simplistic questionnaire, processing time for the data is also very limited.

1.13 Another important characteristic of ETS is that they are a unique source of information about agents' expectations on relevant economic outcomes (production, demand, households' income). They also allow to collect information on economic phenomena for which there is no or insufficient quantitative data; for example constraints in production or inventories behaviour. Furthermore, tendency surveys are not subject to revisions. For all these reasons, tendency surveys are functioning very well as a complement for the quantitative data.

1.14 Referring to CTS, additional considerations have to be made. A survey aimed at measuring the so-called "Sentiment" of consumers was first introduced in the United States (US) as far back as 1946. Nowadays, CTS are carried out in at least forty-five countries worldwide and their outcomes are both widely used in the business and financial press and analysed by economists and policy-makers.

1.15 The original intuition behind the setting of CTS is due to George Katona (1951), which argued that consumers' attitude towards consumption and saving may have a relevant role in explaining consumption patterns. This may be particularly true in periods characterised by exceptional events, when household's decisions may be considered to depend not only on the "ability", but also on the "willingness" to buy.

1.16 According to this view, consumer attitudes cannot be explained only by their reaction to changes in economic variables, but are also influenced by non-quantitative, non-economic factors - such as political crisis or wars – supposed to have an impact on agent's psychological mood. Consequently, the "willingness to buy" may be an important and independent explanatory factor for spending, especially for discretionary purchases (such as durable goods), and in proximity of turning points.

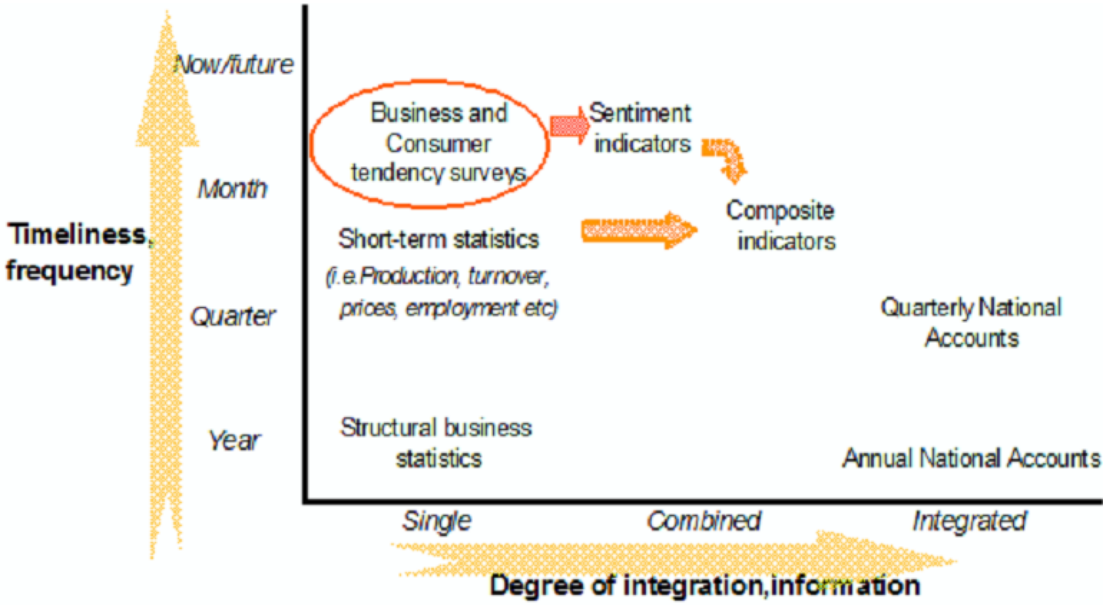
1.17 Since the pioneering works of Katona, the role of psychological motives has been widely debated in the literature; indeed, many different alternative interpretations of the sentiment variable have been advanced (see among others Carroll et al., 1994; Campbell and Mankiw, 1991; Acemoglu and Scott, 1994; Deaton, 1992; Sommer, 2001; Carroll, 2004; Souleles, 2004).

1.18 On the opposite side of the theoretical debate, Sentiment has also however been considered as a mere cyclical indicator, without a proper theoretical role in explaining consumption patterns. This position has been held by many authors, in the spirit of the initial skeptical position taken by the FED (1955).

1.19 The system of available economic statistics and the problem of availability of timely information can be described by a graph containing the time dimension and the degree of integration and information as shown in Figure F1.1. The matrix shows at a glance the position of the various statistics and the interrelationships between them.

**Figure F1.1**

**The system of economic statistics and the need for more timely information**




**Source: based on the concept of statistical matrix discussed by Algera,(2005)**

1.20 All economic indicators can be arranged in this graph and the yellow arrows show the trade-off between more timeliness (wish for faster release of data) and more comprehensive information (wish for more detailed and integrated data). Short-term indicators cover normally one aspect (i.e. turnover, prices or employment) and are released on a monthly or quarterly basis. There is some delay in publication with respect to the end of the reference period. The structural business statistics contain more detailed and reliable information on the variables, but they are only available on an annual base and are usually published over a year after the reference year. On the right hand of the figure, the more integrated and comprehensive information on the economy is described in the quarterly and annual national accounts. Although very detailed, there is a relatively late publication after the reference period.

1.21 So, reading from top to bottom, the reported results become more reliable and detailed, but they take longer to be published. Reading from left to right, the data become more comprehensive and reliable, but again as a rule they are published later. Moreover, as a consequence, the quality of statistics can be assessed in two ways, one in terms of predictive power (the 'columns' of the statistical matrix) and the other in terms of consistency (the 'rows' of the statistical matrix).

1.22 The ETS can contribute in solving the timeliness problem. This data is available either ahead of the end or very shortly after the reference period and they also provide information about expectations for the near future. Furthermore, these information can be used to calculate "confidence" indicators which give some indication for sectors, consumers and the whole economy. And, by combining them with other short-term statistics, in composite indicators even more information is constructed. A well-established indicators system, which uses data from various sources including ETS is the system of Composite Leading Indicators (CLIs) of the OECD.

## C. On the use of ETS

1.23 ETS results are used by various groups. Almost all people and institutions who are involved in business cycle analysis are potential customers of ETS data. Politicians, ministries, central banks, commercial banks, financial markets analysts, associations etc. are potential customers for the survey results. However, also ordinary people are more or less interested in the business cycle development because unemployment, wages, interest rates and so on fluctuate with the business cycle. That is why often the media demands and distributes ETS results.

1.24 Businesses themselves are often interested in BTS results at a sectorial or industry branch level. The information helps them to compare their own position with the sector average. Detailed data about customer sectors or suppliers sectors can help to optimize the own production processes. There is even interest on the survey data on the micro-level: scientists use anonymized micro-data for research.

1.25 So there are various interest groups demanding results from ETS. There is demand for aggregated results, for sectorial results and even for the micro-data. The heterogeneity of user groups means also that the survey organization should develop a publication strategy, which takes into consideration the needs of the different user groups. Media are often interested in brief and timely press releases. Policy makers need more in-depth reports and businesses are interested in

results by industries. Researches usually do not need very timely results. However, when researchers demand micro-data a strategy regarding data access and anonymization is needed.

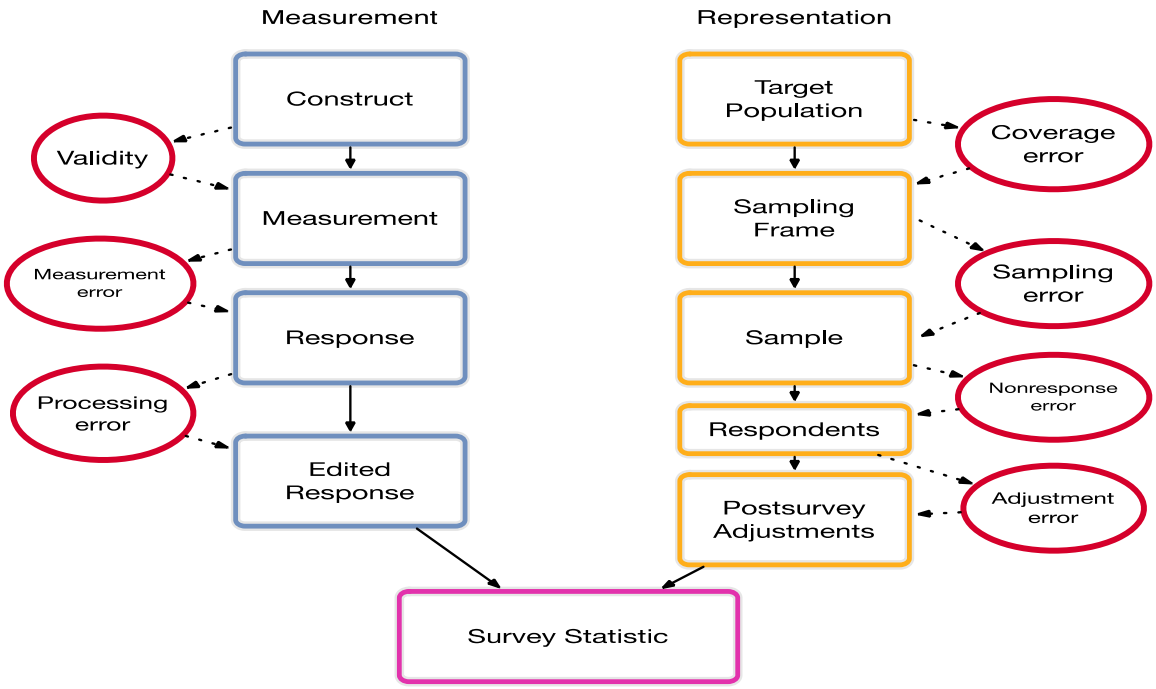
## D. Survey process

1.26 Economic observers usually use time series resulting from regular ETS for their analyses. With judgment and with the help of time series and econometric methods people analyse the current economic situation and try to calculate forecast based, among others, on ETS results. However, behind the indicators which are calculated from regular surveys, stand common sample surveys. As such ETS go through the usual survey process and are prone to usual survey errors.

1.27 Figure F1.2 shows the survey process and common sources of errors as described by Groves et al. (2009). The authors distinguish between the measurement aspect and the representation aspect of a survey. The measurement dimension deals with the question: what is this survey about? The representational dimension focuses on: who is the survey about?

**Figure F1.2**

**Measurement and representation aspects of sample surveys and the possible sources of error**




**Source: Groves et al., (2009)**

1.28 ETS try to capture the economic situation of a country, usually the business cycle development. In order to measure the economic situation various questions were developed. For that purpose a core set of questions emerged over time, which are used successfully in many countries. The questions and the pre-defined response categories are chosen in a way that respondents should be able to answer relatively quick and out of their memories. No or only little calculations or information in accounts and other sources should be necessary for answering the questions. Since the questions and the response scales proposed in this handbook are already successfully used in many countries for a long time the errors in measurement process should be not too large. Sometimes there might be reasons to adapt the standard approach because of cultural, social or other reasons. These adaptions should try to reduce errors in the measurement process.

1.29 Also for the representation dimension of the ETS recommendations are made in this handbook. Representation aspects require usually some adaptions to local circumstances. What is a suitable target population to get survey results, which draw a clear picture about the economic situation in a country? Is there a good database of possible sampling units, which can be used to draw a sample? How could this be sampled? Are the respondents willing to answer? How can response rates be increased? Which survey modes should be chosen – mail, e-mail, internet, phone interview, personal interviews – to obtain high response rates? How should the answers be adjusted or processed after the survey is completed? Business and consumer tendency surveys usually use specific weighting and aggregation schemes. In the following chapters, best-practice for BTS and CTS process is described.

**E. Outline of the handbook**

1.30 This handbook discusses the various aspects of conducting ETS. It also gives hints for the use of the survey results. Chapter 1 defines the scope of the handbook and introduces the main characteristics of the ETS.

1.31 Chapter 2 describes the goal of ETS and contains the main definitions.

1.32 The questionnaire design is discussed in Chapter 3. More specifically, principles guiding the construction and selection of questions are discussed as well as the choice of response scales. Pre-testing of the questionnaire is illustrated, too.

1.33 Chapter 4 focuses on important steps in the entire process of ETS: from where and how to select a subset of units upon (i.e. the sample): survey frame and sample designs suitable for ETS are

illustrated. Moreover, the selection methods of units from survey frame, the allocation criteria and the sample size are discussed.

1.34 Chapter 5 presents considerations regarding the estimation procedure and the accuracy. In BTS and CTS estimation processes, weighting schemes are usually used. Referring BTS and CTS, the goal of the weighting procedures and the choice of specific weights are described in detail. Accuracy is another key word which is discussed in this chapter. Sampling error and its calculation are illustrated.

1.35 Chapter 6 deals with the data collection process. Identification of contact person, survey modes, reminder policy and training of interviewers are topics discussed within this chapter.

1.36 In chapter 7 non-sampling errors and the possible approaches to reduce errors are analysed; in particular coverage, measurement, processing and non-responses (total and partial) errors are discussed.

1.37 Qualitative responses have to be quantified to make them easier to interpret and to obtain time series from different survey waves. Different approaches regarding the process of quantification into a time series are described in Chapter 8. In addition, various standard methods for seasonal de-composition of time series and widely adopted solutions for ETS time series are discussed in this chapter.

1.38 The survey results are used by the public only when they are published in a suitable way. Data dissemination and publication are the topics of Chapter 9. Recommendations regarding metadata and output tables are made. Strategies of publication are also discussed.

1.39 Chapter 10 then presents some elaborated uses of ETS results. The most common business cycle composite indicators are presented and examples of forecasting applications, turning point detection and micro-econometric approaches are given.

## References

Algera, S. (2005). *Statistical benchmarking: the Dutch case*. Paper presented at the Eurostat workshop on Frontiers in Benchmarking Techniques and their Application to Official Statistics, Luxembourg.

Acemoglu, D. and Scott, A. (1994). *Consumer Confidence and Rational Expectations: Are Agent's Beliefs Consistent with the Theory?* The Economic Journal, vol. 104, pp. 1-19.

Campbell, J.Y., Mankiw, N.G. (1991). *The response of consumption to Income. A cross-country Investigation.* European Economic Review, vol. 35.

Carroll, C.D., Fuhrer, J.C. and Wilcox, D.W. (1994). *Does Consumer Sentiment Forecast Household Spending? If so, Why?* American Economic Review, vol.84, n°5, pp.1397-1408.

Carroll, C.D. (2004). *The Epidemiology of Macroeconomics Expectations, in The Economy as an Evolving Complex System III*. ed. by Larry Blume and Steven Durlauf, Oxford University Press.

Deaton, A.S. (1992). *Understanding Consumption*. Oxford University Press, New York.

European Commission, Directorate-General for Economic and Financial Affairs (2014), *"The Joint Harmonised EU Programme of Business and Consumer Surveys"*, user guide March, available at http://ec.europa.eu/economy_finance/db_indicators/surveys/method_guides/index_en.htm.

Federal Reserve Consultant Committee on Consumer Survey Statistics (1955), *Smithies Committee Report*. Reports of the Federal Reserve Consultant Committee on Consumer Survey Statistics, Hearings of the Subcommittee on Economics Statistics of the Joint Committee on the Economic Report. 84th US Congress.Groves.

R.M., Fowler, F.J., Couper, M.P., Lepkowski, J.M., Singer, E., Tourangeau, R. (2009). *Survey methodology.* Wiley, New Jersey.

Katona, G.(1951). *Psychological Analysis of Economic Behaviour*. New York, Mac Graw Hill.

OECD (2003). *Business Tendency Surveys: A handbook.* OECD Publications, Paris, ISBN 92-64-19894-6. (http://www.oecd.org/std/leading-indicators/31837055.pdf

Sommer, M. (2002). *Habits, Sentiment and Predictable Income in the Dynamics of Aggregate Consumption.* Johns Hopkins University Working Paper, n°458.

Souleles, N. (2004). *Expectations, Heterogenous Forecast Errors and Consumption: Micro Evidence from the Michigan Consumer Sentiment Surveys.* Journal of Money, Credit and Banking, vol.36, n°1, pp.39-72.

# CHAPTER 2

## Constructing ETS: scope

### A. Introduction

2.1 The first step in the construction of an ETS is to define its scope in terms of sector and geographical coverage, variables to be inquired, type of the questions and frequency of the survey (see table T2.1).

**Table T2.1**

**Scope of ETS**

| | | |
|---|---|---|
| *Who* | Business or consumers? Which sectors to survey, which firms, which households? Fixed sample or rotating sample? | **Sample design** (chapter 4) |
| *Where* | Geographical coverage? | |
| *What* | What variables covered? What elements of the economy? | **Questionnaire design** (chapter 3) |
| *How* | Structure of the answering scheme, qualitative or quantitative? | |
| *When* | Frequency of the survey? Monthly, quarterly or (bi) annually? | |

2.2 ETS can be conducted among businesses or consumers. Both can be used to analyse developments in economic activity. Businesses give information on the production side of the economy. Consumers give information from a demand perspective. After this selection, further steps include choices on which sectors and firms or which types of households to cover in the survey and what kind of panel will be used.

2.3 Another element is to determine whether geographical factors play a role in economic development. Businesses can be located in different parts of a country, each with specific economic activities. Consumers can be located in more wealthy parts (higher income and employment) or less wealthy parts. If regional differences are considered substantial then this must be included in constructing the etc.

2.4 The third element is a selection of desired variables. Which questions are needed to give insight in economic development? What target indicators have to be monitored? The questions can be backward or forward looking, assessments of the current situation or expectations for the near future. There is some friction between the number of variables asked, the burden on enterprises and

consumer in completing the questionnaire, harmonization with international standards and country specific conditions.

2.5 To make it possible to collect data in a very short period, the answering scheme must be very easy and intuitive. Normally, questionnaires have a fixed number of pre-defined qualitative answering schemes. But in some cases, a quantitative indication can be necessary.

2.6 Finally, the frequency of the survey must be determined. Ideally, ETS are conducted on a monthly basis. However, due to the burden on the respondents, it might be necessary to have different sets of questions with a different frequency. For some variables, it can be sufficient to ask for them quarterly or even (bi) annually.

Because of the different nature of BTS and CTS, they will be treated separately in the remainder of this handbook.

## B. The scope of BTS

### (i) Sector coverage

2.7 BTS can be used to analyse developments at sector or industry level of an economy. They are also used to obtain indicators for the situation of an economy as a whole. To get a good overview about economic development, the sector coverage of the BTS has to be chosen according to some guiding principles. Usually not all sectors of an economy are captured by BTS. This reflects that this type of surveys is no substitute to national accounts. BTS aim to give timely signals about the cyclical development of an economy.

2.8 When constructing a BTS, determine which sectors to cover is the first essential step. Figure F2.1 indicates factors, on which the selection of sectors should be based.

**Figure F2.1**

**Factors relevant for sector selection**

2.9 The size of a sector or an industry, measured for example by value added or employed people, in combination with its cyclicality are the most important decision criteria. Some sectors do not show much variation over the business cycle. These sectors stabilise the economy but from them no strong signals about the business cycle come. Another criteria is whether the sector is controlled by the public, is semi-private or privately organized. Sectors where the development is determined by the public are often not of interest when the aim is to get timely signals about the cyclical development of an economy. Some sectors move rather anti-cyclically, and this can be disrupting especially when cumulating the different surveys to a country-wide sentiment indicator.

2.10 Other characteristics which may be important for the decision of inclusion are the possibility to get high response rates and reliable responses. Is a sector dominated by micro-firms? Is the sector dominated by informal activities or black work? Also weather and seasonal influences must be considered, before selecting a sector.

2.11 Core sectors, which are covered by BTS in many countries, are manufacturing, construction, retail trade and a selection of further services. The harmonized EU programme of DG ECFIN consists of BTS in these sectors (and the CTS).

2.12 Manufacturing is still important to observe, although in most countries not the largest sector, neither by value added nor by employment. Some industry branches are very cyclical and with the expenditure approach in national accounts in mind, demand of machinery and equipment is an important cyclical component of fixed investment. Fluctuations in machinery and equipment expenditures are closely related to the manufacturing sector.

2.13 Another important part of fixed investment is construction investment. So it is very reasonable to cover also the construction sector. In many emerging countries construction is an important driver of economic development. Beyond that, this sector has often been involved in economic crises in different countries, emerging or industrialized ones.

2.14 Retail trade is usually covered because retailing is a quite volatile part of consumption and consumption contributes in many countries a large proportion to Gross Domestic Product (GDP). Especially spending on durable consumer goods is closely related to the cyclical movement of the business cycle. The large weight of consumption also explains why in many countries not only retailers are surveyed but also the other side of the market, the consumers. Retailer surveys in addition with CTS should give a timely and reliable picture about the cyclical development of consumption.

2.15 Although manufacturing is often regarded as one of the important cycle maker the value added of the service sector is in many countries much larger. The service sector is a quite heterogeneous sector, however. There are services to the businesses, services to individuals and services to the public. Some services are linked to fixed investment demand and some are related to consumption (private or public). Some services are market driven - e.g. by Information Technology (IT) - and some are controlled more or less by the public (e.g. health services). Some service sectors might be dominated by a few companies or even monopolies (e.g. telecommunication or parts of transportation) and some are characterized by micro-businesses (e.g. personal services).

2.16 This heterogeneity is the reason why often only a selection of services is covered by BTS. The selection is more country specific than in the previously mentioned sectors of manufacturing, construction and retailing. It depends on the specific market structure. Which services are market oriented and cyclical? Which services are important for a specific country? Furthermore, the heterogeneity also has some impact on the selection of variables covered. For example, businesses like hotels or restaurants which are in the same questionnaire as businesses in transport might have some problems with some of the variables asked.

2.17 Another problem area within the service sector is the treatment of the financial services. Although very important for most developed countries, there are some objections against including them in the survey for services. The number of businesses in a country is often limited and also a part of the selected variables are not applicable for the financial services. In Europe, DG ECFIN has countered these drawbacks to conduct a special survey among financial services, on a European sample, instead on a sample per country.

2.18 Some institutions survey not only retailers but also wholesale traders. Wholesaling has various distribution functions. It distributes intermediate products to manufactures and consumer goods to retailers. So wholesale is connected to the producing sector and to consumption. There is even a further function of wholesalers, they import and export goods. So wholesaling is a sensible sector to cover in countries with large export activity. However, wholesaling could also be a sensible indicator for countries which depend to a large extend on imports. This could be of especial interest for example in developing countries with small manufacturing sectors.

2.19 Also country specific could be the inclusion of further sectors like mining or agriculture. The decision about the inclusion should be guided by the factors shown already in Figure F2.1, which include the characteristics size, cyclicity and market structure. However, they should be

included only when they are at least important in terms of GDP. Economic activity in these sectors is particularly influenced by weather conditions, seasonality and anti-cyclical behaviour. If they do not contribute substantially to GDP their coverage might not be necessary for the assessment of the general economic situation.

2.20 When a sector is selected, the level of detail for publication must be determined. Common strata in BTS are type of activity (usually expressed in an adaption of a ISIC[2] classification) and firm size (often measured by employed people or turnover). In some countries, the geographical dimension is covered as well (see below).

**(ii) Regional Coverage**

2.21 To define the intended sectorial coverage of a BTS is an essential step in the set-up of such a survey. However, one should also take care about further dimensions of interest. A question one should raise is the question about the desired regional information level. Is an analysis of local or regional entities intended? If yes, this has consequences for the sampling procedure and the reporting units. For sampling, the question is whether the local dimension is important enough to introduce it as a strata. Location would then be a further strata. However, even if location is not used as strata and the local distribution is more or less left to the random process, local statistical evaluation might be of interest.

2.22 To calculate local statistics one has to have the information about the location of a firm. This is often not very problematic because one needs the contact details anyway. The problem is what to do when firms are located at more than one local entity. In that case one has already at the beginning of the survey implementation to decide how to proceed.

**(iii) Survey units**

2.23 In conceiving BTS it is useful to distinguish three units. The reporting unit is the part of an enterprise for which data are collected. The sampling unit is contained in the survey frame and is used for sample selection. And the response unit is the unit to which the questionnaire is sent and gives the answers about the reporting unit.

[2] The International Standard Industrial Classification (ISIC) is a United Nations system for classifying economic data.

**Table T2.2**

**Breakdown of enterprises**

| *Enterprise* | *Location 1* | *Location 2* | *Location 3* |
|---|---|---|---|
| *Product 1* | *Establishment* | *Establishment* | *Establishment* |
| *Product 2* | *Establishment* | *Establishment* | *Establishment* |
| *Product 3* | *Establishment* | *Establishment* | *Establishment* |

2.24 Why is such a separation sensible? As Table T2.2 shows enterprises can be split regarding location and products. Some enterprises produce more than one product and are located at more than one location. The local production units are here called establishments. Regarding the two dimension there are rough or approximate procedures and there are more elaborated ones. If an enterprise is active in more than one ISIC class the approximate method would be to count the response for that class where the enterprise has its main focus, measured for example by turnover. The more precise procedure would be to ask the enterprise to fill out one questionnaire for each activity. In that case there are several reporting units for one enterprise and there could be more than one response unit. However, it might also be that one response unit gives answers for more than one reporting unit.

2.25 Regarding the treatment of local units the intended level of analysis is important. If location is important a rough method would be to use the firm's main seat to localize it. Another approximate method is to "distribute" the response to various locations according to size information of local units, e.g. number of local employees. The most precise procedure is to survey local establishments. If local analyses are important, the surveying of local establishments is the best method. However, in practice one is often forced to use a mixture of the exact and some approximate procedure because usually not all enterprises are willing or even able to give answers for establishments. It may also be the case that local units do not have enough information to answer questions, especially regarding assessments and plans. Sometimes local units are not allowed to give information because only the central office is allowed to do this.

**(iv) Basic principles of questions**

2.26 BTS are conducted in various countries since many years ago. Over time some standards regarding survey questions and response scales emerged. Details about the questions are discussed later in this handbook. The selection of questions is based on some guiding principles:

- Since BTS aim to capture business cycle developments timely questions are selected which measure business activity at an early stage.
- The question topic should be sensitive to changes in the economic environment.
- BTS focuses not only on hard facts but also on assessments, expectations and plans.

2.27 In addition to business cycle relevant topics, extra questions about bottlenecks could be asked. This type of questions ask for factors hindering business activity and are often asked less frequently than the usual questions. Especially in developing and emerging countries this type of questions could be interesting for policy recommendations or for measuring reform progress.

2.28 Another aspect, which is important to achieve cooperation from respondents, is that the questions should be relevant also in the view of managers of enterprises and not only from a business cycle theory perspective. This is important since BTS are mostly voluntary surveys and are panel surveys. Respondents are asked regularly. They only will answer regularly when the survey questions (and results) are relevant for them.

2.29 Standards emerged not only for drafting and selecting questions but also for response scales. Most of the responses are measured on a qualitative scale. Some usual pre-defined answer categories are good/usual/bad or improve/remain unchanged/deteriorate.

2.30 Regarding the sampling procedure qualitative information usually needs less sophisticated sampling plans than quantitative information. In addition the required sample size is smaller with qualitative data. Indeed this means that qualitative data is not as precise as quantitative data of high quality. That is why qualitative data is often regarded as a timely supplement to quantitative data, not a substitute.

2.31 However, regarding some topics there is for the survey planner no choice between a qualitative and a quantitative scale. For example when it is asked about assessments. It would be almost impossible to assess the stock of finished products on a continuous quantitative scale. Overall, the experience with BTS shows that they give timely and valuable information about the economic course. This information is demanded by various user groups both in countries with a sophisticated system of national accounts and other official statistics and in countries with less sophisticated national statistical systems.

**(v) Survey frequency**

2.32 Behind the decision about survey frequency is a trade-off between the workload for respondent and for the survey institution and the need for timely signals about the economic course. To monitor the economic development indicators should have a monthly or at most quarterly frequency. Biannual or annual data is more relevant for structural analyses than for monitoring the business cycle. For policy makers the signals of annual data are too late. Policy measures like stimulus packages need time for implementation, so that timely signals are important. Thus from a user perspective a monthly survey is optimal. However, a monthly frequency must be accepted by a sufficient number of respondents. And in addition the survey institution needs the resources to implement and conduct a monthly survey. So there is a weighting-up process between monthly and quarterly survey frequency. If possible the aim should be – at least in the long run – a monthly survey frequency.

### C. The scope of CTS

2.33 As for BTS, the first step in the construction of CTS is to define its scope in terms of coverage, variables to be inquired, type of the questions and frequency of the survey (see table T2.3).

**Table T2.3**

**Scope of CTS**

| | | |
|---|---|---|
| *Who* | Individual consumer or households? Fixed sample or rotating sample? | **Sample design** (chapter 4) |
| *Where* | Geographical coverage? | |
| *What* | What variables covered? What elements of the economy? What is the time horizon of the questions? | **Questionnaire design** (chapter 3) |
| *How* | Structure of the answering scheme, qualitative or quantitative? | |
| *When* | Frequency of the survey? Monthly, quarterly or (bi) annually? | |

#### (i) Target population and sampling and reporting unit

2.34 The target population for CTS is represented by a country's total adult population, from which to extract a sample of representative individual consumers, selected on the basis of its socio-economic and demographic characteristics.

2.35 The sampling unit is usually an individual selected within a given household; the reporting unit is generally coincident with the sampling unit.

2.36 Sampling variables generally include the municipality or more generally the geographical area of residence, gender, the employment status and information about income. The sample is usually extracted from as a comprehensive as possible list, ideally based upon vital statistics registers. More detailed recommendation on sampling practices are contained in chapter 4.

**(ii) Principle of the questions**

2.37 Any analysis on the role of Sentiment in explaining consumption patterns starts from the setting out of an appropriate questionnaire, designed to capture consumers' attitudes and expectations towards saving and spending.

2.38 Survey questionnaires usually contain a fairly large number of qualitative questions (15, in the cases of the EU-Wide Harmonised Project) generally allowing five possible replies arranged on a Linkert scale ranging from the extremely positive to the extremely negative. The proper cyclical part of the questionnaire is generally preceded by a section aimed at gathering structural information on the respondent and her household.

2.39 Structural information usually include data about demographic characteristics of the respondent and the household, together with socio-economic characteristics of the respondent.

2.40 Cyclical questions usually ask information about the economic situation of the household and that of the country, saving and buying intentions, with particular emphasis on durable goods.

2.41 In terms of time horizon, questions are usually referred either to the situation of the last 12 months or to that of 12 months ahead.

2.42 Qualitative questions generally allow five possible answers, ranging from extremely favourable to extremely unfavourable. For quantitative questions, replies are usually bounded and confirmation is asked in case the reply mentions extreme values. For more detailed recommendation on the building up of the CTS questionnaire, see chapter 3 below.

**(iii) Survey mode and frequency**

2.43 CTS are usually performed by telephone. Fixed lines are normally used; more recently in some cases mobile phones are used to. The use of the internet is generally limited in the case of CTS; in fact, in many countries the diffusion of the internet is still not fully homogeneous across the population, and hence the use of the internet mode may bias survey results. Moreover, internet surveys imply a stronger commitment to reply from the respondents, which is often more difficult to

achieve for the consumers than for the firms, since for the consumers the relevance of survey information is less immediate than for firms.

2.44 CTS are usually performed on a monthly basis; some of the questions (more specifically, those on the intention to buy durables) are often administered on a quarterly basis.

## References

European Commission, Directorate-General for Economic and Financial Affairs (2014), *"The Joint Harmonised EU Programme of Business and Consumer Surveys"*, user guide March, available at http://ec.europa.eu/economy_finance/db_indicators/surveys/method_guides/index_en.htm.

OECD (2003). *Business Tendency Surveys: A handbook.* OECD Publications, Paris, ISBN 92-64-19894-6. (http://www.oecd.org/std/leading-indicators/31837055.pdf).

# CHAPTER 3

# The questionnaire design

## A. Introduction

3.1 Business cycles are an important feature of the economies of market-oriented industrialized countries. Statistical series derived from BTS are particularly suitable for monitoring and forecasting business cycles. Indeed, as highlighted in chapter 1, data from BTS are released soon after the end of the reference month, making it easier for policy makers and analysts to analyse the business situation in real-time. Cyclical profiles of the series are in many cases easy to detect because they contain no trend. Usually the series are seasonally adjusted, at least to some extent, by respondents, adding smoothness to the series. This and the fact that they usually do not need revisions, facilitates their use in forecasting and, in particular, in predicting turning points in the business cycle on a short horizon.

3.2 In order to be able to monitor business cycles and to compare them between countries, the information collected in the business and consumer tendency surveys needs to be precise and standardised at an international level. It is therefore relevant to define a frame of internationally standardised sets of questions used as a reference for participating countries. The EU harmonised programme is a well-established example of such a framework. It is especially relevant for industrialised countries, even if they do not belong to the EU. Some adaptations to this framework may also be considered for developing countries.

3.3 This chapter will explore how to design business and consumer tendency questionnaires in a standardized way. Starting from the formulation of questions and the methods of measuring the responses, it aims to provide a framework for the standardization of questions in their time dimension, scale and period covered across sectors of economic activity (section A). Also pre-testing of the questionnaire is discussed here. A set of core variables for different frequencies (monthly, quarterly and annual) is presented in section B, for both business and consumer surveys.

## B. General considerations

### (i) Business and consumer tendency surveys: main differences

3.4 BTS substantially differ from CTS by the type of information collected and more broadly by their scope.

3.5 *BTS* are designed to capture information about the activity of an enterprise, following the entire business cycle from orders, production, stocks and finally sales[3]. The accuracy in terms of measuring the production process is rooted to the enterprise itself, and given that the targeted respondents are usually business managers, the answers are expected to be precise and closer to a quantitative assessment rather than to a qualitative perception of the production. Questions also refer to a relatively short time span (either evaluating the current situation or assessing past and future evolutions), adding soundness to business managers' responses as they can easily have access to the information required for the most recent past and provide an accurate estimation for the close future.

3.6 *CTS* consist of questions dealing with the individual financial situation of households as well as their perception on the general economic situation in their country. Hence, they have two main purposes: to cover the demand-side of the business cycle and to assess consumer confidence. In the EU harmonised programme, a set of "micro-questions" focuses on households' financial situation, past, current and future. Consumers are also asked if they are likely to make durable consumption expenditures in the present or in the future[4]. Moreover, a set of "macro-questions" is devoted to the perception of the economic situation in the country, such as the evolution of consumer prices and unemployment. The questionnaire could lead to some confusion if questions were poorly formulated with no clear distinction between the household level and the country level as consumers may easily misunderstand the target of the questions.

### (ii) Formulation of questions

[3] In the EU harmonised programme, questions on sales are only included in the retail-trade and in the services surveys. However, a slightly different harmonised framework may be suitable for developing countries.

[4] In the monthly EU survey, consumers shall assess the appropriateness of major purchases at the moment and indicate whether the amount of money spent on major purchases over the next 12 months will be higher or lower than in the preceding period. In the EU quarterly survey, they are inquired about the likelihood of buying a car or buying/building a home over the next 12 months.

3.7 Reducing the response burden of a survey often allows increasing the response rate. Therefore, questionnaires should be designed to be self-contained in the instructions, straightforward in the compilation and easily accessible. Their content should also be relevant to business managers and consumers in order to motivate their participation. Few key variables should be included in the form, especially if the survey is to be conducted on a monthly basis. It should be clear to respondents that confidentiality on the information supplied is preserved as only aggregated results will be published. Disaggregated data at the micro level may be used for research purposes, for example to measure the dispersion in individual responses and propose means of correction for potential bias, while preserving confidentiality.

3.8 Departures from the standard questionnaires in terms of concepts used, reference periods or type of questions, should be kept to a minimum. However, it is possible to include additional questions that may pertain to *ad hoc* topics, for which either a separate questionnaire or an additional set of questions to an existing survey is envisaged. In the case of the Joint Harmonised EU Programme of Business and Consumer Surveys, each month the EC uses a very short questionnaire for the regular BTS while adding additional questions every quarter or half-year to monitor other aspects of the economy for the participating countries. The quarterly questions are always the same and the only reason not to include them on a monthly basis is to alleviate the response burden on participants[5].

3.9 An important aspect to consider when designing a standardised questionnaire is related to the *wording of questions* in national languages. Special care should be devoted to the translation phase of questions. Rewording or rephrasing questions in order to adapt the standardised questionnaires to various national languages should be limited to exceptional cases, as translations that deviate from the literal meaning of the original questions are prone to introduce errors and distortions in the survey. The recommendation of this manual is to keep to a minimum the number of departures from literal translations, and when unavoidable, it is suggested to preserve the original meaning in the reworded questions.

3.10 Questionnaires should also cover the *same reference period* in order to facilitate cross-country comparisons (see subparagraph A(v)). Moreover, data should be collected at the same time in all countries. In the EU, fieldwork for monthly surveys is generally performed in the first two

[5] See the user guide published by the EC for more details: http://ec.europa.eu/economy_finance/db_indicators/surveys/documents/userguide_en.pdf.

weeks of each month, while quarterly surveys are carried out in the first half of each quarter (January, April, July and October).

**(iii) Choosing subjects and sectors**

3.11 *The choice of information* collected in business and consumer tendency surveys questionnaires is an exercise that must be performed bearing in mind the needs of final users. Variables should be chosen for their relevance to monitor and forecast business cycles. The topic should also be relevant to both business managers and consumers, as a mean of incentive for their continuous participation to the survey.

3.12 In *BTS*, different types of information may be interesting to collect:

a) Expectations of future sales/production/business activity[6], possibly also order books;
b) past sales/production/business activity;
c) Stocks of finished goods;
d) Selling prices;
e) Employment;
f) Capacity to export: competitiveness and export order books;
g) Capacity utilization, factors limiting production.

3.13 The EU harmonised surveys draw in each of these seven categories, but not necessarily each month and for each sector of the economy. In developing countries, the choice of variables may be specific. For instance, companies may not have any order books in certain countries and questions related to them become irrelevant in this case. Specific BTS are usually conducted for the main sectors of the economy. In the EU harmonised programme, five sectors are considered: industry, retail trade, construction, financial services and other services. Of course, different divisions or sub-divisions could be implemented. It depends on the budget allocated to tendency surveys and on the structure of the economy. In developing countries, a specific survey for agriculture could be considered, for instance, depending on the contribution of this sector to GDP. In developed countries, it could also be useful to divide the services sector into more homogeneous sub-sectors as transport, business services and recreation services are quite different from each other, for instance.

3.14 In *CTS*, two types of questions are generally considered:

a) micro - (fact based) questions

[6] Depending on the sector to which they belong, firms may be asked about their sales, their production or their business activity. All these terms refer to output.

b) macro - (sentiment based) questions.

Questions regarding the financial situation of households, their intention or capacity to save, their intention to buy durable goods, both in the past and in the future, belong to the first group. Those regarding the general economic situation in the country, the evolution of unemployment and consumer prices, in the past and in the future, belong to the second group. As for BTS, not all questions are necessarily asked on a monthly basis.

**(iv) Measurement scale**

3.15 The majority of questions in *BTS* data are formulated on a three-option ordinal scale. Typical possible answers are listed below:

- Up / unchanged / down
- Above normal / normal / below normal
- Too large / adequate / too small
- Increase / stable / decrease
- Better / same / worse
- Easy / normal / tight
- Good / satisfactory / bad.

The last set of options is only used in questions assessing the present, as the evaluation scale (good, satisfactory or bad) requires no comparison to the past or the future.

3.16 There may be exceptions in BTS where questions are not formulated around a three point scale. For instance, questions on capacity utilisation should refer to a percentage of full capacity utilisation; questions on the number of months of production assured by the orders at hand should, of course, refer to a number of months; questions asking which factors are limiting production, activity or investment should refer to a list of factors to choose from. Finally, there are some questions on the structure of investment, as well as quantitative questions in the investment survey.

3.17 In *CTS*, questions are generally formulated around a five option scale. Typical possible answers are listed below:

- A lot better / much more / increase sharply
- A little better / a little more / increase slightly
- The same / about the same / remain the same
- A little worse / a little less / fall slightly

- A lot worse / much less / fall sharply

Questions may also be formulated around a four option scale. In this case, the neutral answer is generally removed from the available options.

3.18 As the information collected in BTS is mainly qualitative with replies framed in a Likert scale type of measure (three to five options), one needs to find a way to interpret it. Percentages obtained from the multiple-choice questions are usually converted into a single number by using net balances or diffusion indices. Computational details of net balances and diffusion indices are provided in chapter 8.

**(v) Reference period**

3.19 In order to achieve comparability over time, across countries and across questions, it is desirable to maintain in the harmonised questionnaire the same time dimension as well as the same measurement scale across questions and across countries. Questions used in ETS can be classified regarding to their time dimension, as they can be addressed to the recent past, the current situation or the near future.

3.20 In monthly and quarterly *BTS*, a time span of three months may be considered as a benchmark in order to assess past and future evolutions as well as the present situation. This advice is consistent with the EU harmonised programme. However, when institutes supplement the survey questions from the EU BCS questionnaire with additional, self-designed, ones, they sometimes use deviating time-horizons. An example is the question by the Information and Forschung (Research) Institute (IFO) on business expectations in Germany for the next six months. Answers to this question are then used to compute the business climate in Germany[7]. The time span may also be different for (semi-) annual surveys. For instance, following the EU guidelines for the semi-annual investment survey, manufacturing firms are asked about the evolution of their investment in the previous year and about their expectations for the year to come.

3.21 Other considerations should be taken into account for *CTS*. Unlike business managers, consumers do not think in month-to-month or quarter-to-quarter terms and generally do not keep accounts of their past decisions. Therefore, the most natural benchmark for consumers in order to gauge their present or future situation is an annual (year-on-year) comparison. Another advantage of

[7] For more information on the computation of the IFO business climate, see: http://www.cesifo-group.de/ifoHome/facts/Survey-Results/Business-Climate/Calculating-the-Ifo-Business-Climate.html

a larger time-span is to reduce the volatility of the responses. This advice is consistent with the EU harmonised programme.

**(vi) Seasonality**

3.22 The *seasonal component* in time series corresponds to the regular movements observed in quarterly and monthly time series during a twelve-month period. Examples of these include increases in retail sales data during the Christmas period or the fall in industrial activity during vacation periods. In addition to the effect of seasonal influences, a second type of variation which is also linked to the calendar can be observed. This is the trading day effect. For "flow" data (i.e. data calculated by adding daily figures) the trading day effect arises because of the varying number of such days in a month. For example, a monthly time series of retail sales would be affected by the number of Saturdays in each month. In the case of "stock" data referring to a particular period in the month (for instance the last working day) the calendar effect corresponds to the importance of the day of the week when data are measured. Presenting a time series from which the seasonal movements have been eliminated allows the comparison of data between two months or quarters for which the seasonal pattern is different

3.23 In many industries, managers are aware of the seasonal patterns affecting their business production, sales, stock levels, etc. It is therefore important to *tell the respondents whether or not they should take it into account in their answers*. This information can be given as part of the general instructions that will accompany the questionnaire, but common practice is to repeat the instruction in all questions where seasonality is likely to be important. As a good practice, questions will then start with a phrase such as "*Ignoring seasonal factors,* are stocks of finished goods…?" or "*Excluding seasonal variations,* are sales…?" However, experience in handling BTS data shows that this tends to reduce seasonality but does not eliminate it entirely. The recommendation of this Handbook is that the respondents should be asked to exclude seasonal variations but that all data, after having been converted into net balances or diffusion indices[8], should be systematically tested for seasonality and seasonally adjusted when needed[9].

**(vii) Structural Information on respondents**

3.24 Business and consumer tendency surveys are structured in a way to collect qualitative information regarding the company and the consumer perceptions as well as provide a set of

[8] The reader is referred to chapter 8 for computational details on net balances and diffusion indices.
[9] Seasonal adjustment methods suitable for ETSs series are treated in chapter 8.

*structural information about the company and the household surveyed*. As qualitative information regarding companies and households has been described in the previous paragraphs, this part provides some drives on which type of structural information is required for both surveys.

3.25 To complement the information collected regarding businesses, *BTS* should include the following items:

(a) Sector/ Major economic activity/ Product line
(b) Company name
(c) Company address
(d) Employment size
(e) Name and position of respondent
(f) Contact information (telephone number, facsimile number, email address)

3.26 Similarly, *CTS* should include the following items:

(a) Name of respondent
(b) Name of household head
(c) Address
(d) Respondent's characteristics
    (i) Relationship to the household head
    (ii) Age
        a. 16-29
        b. 30-49
        c. 50-64
        d. 65+
    (iii) Sex
        a. Male
        b. Female
    (iv) Civil status
    (v) Highest educational attainment
        a. Primary
        b. Secondary
        c. Further education

(vi) Working status

a. Full time

b. Part time

(vii) Primary occupation

a. Self-employed professional

b. Self-employed farmer

c. Skilled manual worker

d. Other worker

e. Other occupation

f. Unemployed

g. Inactive

(e) Household characteristics

(i) Household size

(ii) Number of employed persons

(iii) Number of overseas workers[10]

(iv) household income

a. 1st quartile

b. 2nd quartile

c. 3rd quartile

d. 4th quartile

**(viii) Pre-testing of the questionnaire**

3.27 Pre-testing of the questionnaire is a very important part of a questionnaire design. The procedure consists in a systematic check of the data collection tool prior to the implementation on the full scale in order to identify possible problems of survey execution and provide solutions before the official release. Rather than reporting results, pre-testing serves the purpose of controlling for possible problems on the survey implementation, for example it helps assessing whether the questions are relevant and easily understood by the respondents, if instructions are readable, if questions order is appropriate etc. Pre-testing is also useful to determine the duration of interviews and it proves to be a good tool to assess respondents' willingness to participate to the survey.

[10] This may apply to developing countries only.

3.28 Usually if some issues occur in the pre-testing phase, it is likely that similar problems will arise during the full scale administration. Thus pre-testing becomes essential in order to obtain better accuracy and interpretability of the survey results. Pre-testing is administered to a small set of respondents who represent or are drawn from the population of interest. Data collection conditions as well as characteristics of the target population should be as close as possible to the actual survey. As there is no minimum required number of respondents for the pre-testing, the number should be sufficient enough to determine whether there are changes needed to be made to improve the survey instrument.

3.29 Other tools useful to check the robustness of a survey before the actual implementation are the *pre-field* and *the field pre-testing techniques*. Pre-field techniques are generally used during the preliminary stages of questionnaire development. Field techniques, testing questionnaires under operational conditions, include behaviour coding of interviewer / respondent interactions, interviewer debriefings, respondent debriefings, split-sample tests, and the analysis of item non-response rates and response distributions. Some points worth considering when conducting the pre-testing:

- Detailed notes on encountered problems should be included and possible solutions should be identified;
- It is beneficial to pre-test the questionnaire with specialists in question construction, who may be able to identify potential issues otherwise difficult to see in a pre-test with respondents;
- If there are a variety of respondent types, all should be included in the pre-test, and if the questionnaire is designed to be in several languages or dialects, it should be tested in each language or dialect.

Based on the results of the pre-testing, potential modifications that can be made to the survey include: (a) revise the structure of the questionnaire, such as the order or sequence of questions, filters, etc., (b) add more instructions to the questionnaire and interviewer's manual, (c) delete redundant questions / add new ones, and (d) reword or rephrase questions.

3.30 After resolving the problems and issues encountered during the pre-testing, the next step will be to finalize the survey questionnaire ready for implementation on the full scale survey.

**C. An example of harmonised questionnaire**

3.31 The following example draws heavily on the joint harmonised EU programme of business and consumer surveys which is considered as a reference for all developed countries, even outside of the

EU. Questions and sectors considered in the EU programme are marked in bold. For the detailed wording of these questions, the reader should refer to the user guide published by the EC[11]. Other possible questions and sectors (e.g.: agriculture) have been added in order to complete the picture and because they could be interesting for developing countries[12]. Of course, countries wanting to compare their business cycles or to aggregate their business and consumer tendency surveys results should agree on a set of harmonised questions in the following list. As a general rule, it is recommended to introduce at least all questions from the harmonised programme of the EU.

3.32 For all questions, respondents should be requested to abstract from purely seasonal fluctuations. After having been converted into net balances or diffusion indices, answers should also be seasonally adjusted when needed (see subparagraph A(vi)). If a survey has been introduced too recently for traditional seasonal adjustment methods to be implemented, one can temporarily ask for additional comparisons with the year before.

**(i) BTS at a monthly frequency**

3.33 Table T3.1 shows an extensive set of possible questions for a monthly BTS. Concerning the time span for questions referring to the past or to the future, three months can be considered as a benchmark.

3.34 Questions referring to the evolution of production, business activity or sales, depending on the sector, over the last *x* months allow nowcasting economic activity (e.g. GDP) before national accounts figures are released. Those referring to its evolution over the next *x* months allow forecasting economic activity.

3.35 The current level of order books is expected to give an indication on the evolution of production or business activity in the future. Of course, the link between current order books and future production also depends on factors limiting production and the capacity utilisation rate. Similarly, the past level of order books may help to now-cast the evolution of production over the last *x* months. However, practical experience has shown that questions on orders books may be less relevant for developing countries than for advanced economies. In this case, questions on the evolution of production, business activity or sales will be more relevant.

[11] http://ec.europa.eu/economy_finance/db_indicators/surveys/documents/userguide_en.pdf.

[12] Even in the EU, countries typically add questions to the EU harmonised programme. In BTS, possible additional questions may concern the economic situation of the country, the evolution of consumer, producer or export prices, the stock of raw materials and intermediate goods, the flow of new orders (more specific than the level of order books in the harmonised questionnaire), the financial situation of firms and their access to credit.

## Table T3.1

## Monthly BTS

| Sector | Subject | PAST (last *x* months) | CURRENT | FUTURE (next *x* months) |
|---|---|---|---|---|
| **INDUSTRY** | Evolution of production / business activity | **x** | | **x** |
| | Evolution of employment | x | | **x** |
| | Level of Order books | x | **x** | |
| | Level of export order books | x | **x** | |
| | Stock of finished goods | | **x** | |
| | Evolution of selling prices | | | **x** |
| **CONSTRUCTION** | Evolution of business activity | **x** | | x |
| | Evolution of employment | x | | **x** |
| | Level of order books | x | **x** | |
| | Factors limiting production | | **x** | |
| | Evolution of selling prices | | | **x** |
| **RETAIL TRADE** | Evolution of sales / business activity | **x** | | **x** |
| | Evolution of employment | x | | **x** |
| | Evolution of orders placed with suppliers | | | **x** |
| | Stock of finished goods | | **x** | |
| | Evolution of selling prices | | | x |
| **FINANCIAL SERVICES** | Evolution of business activity | **x** | | x |
| | Evolution of employment | **x** | | **x** |
| | Evolution of demand / turnover | **x** | | **x** |
| **OTHER SERVICES** | Evolution of business activity | **x** | | x |
| | Evolution of employment | **x** | | **x** |
| | Evolution of demand / turnover | **x** | | **x** |
| | Evolution of selling prices | | | **x** |
| **WHOLESALE** | Evolution of sales / business activity | x | | x |
| | Evolution of employment | x | | x |

| | | | | |
|---|---|---|---|---|
| | Stock of finished goods | | x | |
| | Evolution of selling prices | | | x |
| **AGRICULTURE** | Evolution of production / business activity | x | | x |
| | Evolution of employment | x | | x |
| | Factors limiting production | | x | |
| | Stock of finished goods | | x | |
| | Evolution of selling prices | | | x |

### (ii) BTS at a quarterly frequency

3.36 Quarterly questions supplementing monthly topics can be added for most sectors. The reason to ask these questions at a different frequency is to reduce the response burden on monthly questionnaires. Some variables typically do not need to be asked for every month. As an example, questions on factors limiting production and assessment of production capacities usually are asked at a quarterly frequency because they are not prone to rapid changes. As for monthly questionnaires, three months may be considered as a benchmark time span for all questions referring to the past or to the future.

**Table T3.2**

**Quarterly BTS**

| Sector | Subject | PAST (last *x* months) | CURRENT | FUTURE (next *x* months) |
|---|---|---|---|---|
| **INDUSTRY** | Evolution of order books | **x** | | x |
| | Evolution of export order books | | | **x** |
| | Months of production assured by current order books | | **x** | |
| | Factors limiting production | | **x** | |
| | Production capacity | | **x** | |
| | Capacity utilisation | | **x** | |

| | | | |
|---|---|---|---|
| | Stock of raw materials | x | |
| | Evolution of the competitive position on the domestic market | **x** | |
| **CONSTRUCTION** | Months of production assured by current order books | **x** | |
| | Capacity utilisation | x | |
| **FINANCIAL SERVICES** | Evolution of operating income | **x** | **x** |
| | Evolution of operating expenses | **x** | **x** |
| | Evolution of profitability | **x** | **x** |
| | Evolution of capital expenditure | **x** | **x** |
| | Evolution of competitive position | **x** | **x** |
| **OTHER SERVICES** | Factors limiting activity | **x** | |

NB: There are no questions asked at a quarterly frequency for the retail trade sector in the EU harmonised programme.

**(iii) BTS at a (semi-) annual frequency**

3.37 Fixed investment amplifies the business cycle and sometimes even drives it. As such, early information about the change of investment intentions is interesting for business cycle analysts. Therefore, the EU harmonised programme includes a semi-annual investment survey for the manufacturing sector asking about changes in investment. Since investment plans are not revised frequently, the EU keeps this questionnaire separate from the monthly survey and instead runs it twice a year. The EU survey mixes quantitative questions on the relative evolution of investment in the last year and in the next year (% change) with qualitative questions on the structure of investment and on factors influencing it (demand, financial resources or expected profits, technical factors, other factors).

3.38 In order to reduce the response burden, this manual only recommends including quantitative questions regarding the evolution of investment, ideally twice a year and in different sectors. In a first survey, at the beginning of year *t* before annual national accounts for year (*t-1*) are released, firms would be asked for the change in investment between year (*t-1*) and (*t-2*) and between year t and (*t-1*). In a second survey, six months after, firms would be asked for the change in investment between years t and (*t-1*) and between years (*t+1*) and *t*. Surveyed sectors should include, at least, industry and other sectors contributing significantly to the evolution of aggregate investment in the country.

**(iv) CTS at monthly and quarterly frequencies**

3.39 Questions in CTS can also be classified regarding the time dimension with the categories recent past, current situation and near future. However, the time span used is usually larger than for BTS. It is in fact common to look at the 12 months back and forward (see subparagraph A(v)).

3.40 The joint harmonised EU programme defines a complete set of questions for CTS. Table T3.3a and T3.3b present these questions which form a core set for these surveys.

**Table T3.3a**

**Monthly CTS**

| Subject | Past (last *x* months) | Current | Future (next *x* months) |
|---|---|---|---|
| Evolution of household financial situation | **x** | | **x** |
| Evolution of the general economic situation in the country | **x** | | **x** |
| Evolution of consumer prices | **x** | | **x** |
| Evolution of the number of unemployed people | | | **x** |
| Right moment to make major purchases | | **x** | |
| Evolution of money spent on major purchases | | | **x** |
| Opportunity to save | | **x** | **x** |
| Statement best describing the household financial situation | | **x** | |

**Table T3.3b**

**Quarterly CTS**

| Subject | Past (last *x* months) | Current | Future (next *x* months) |
|---|---|---|---|
| Likelihood to buy a car | | | **x** |
| Plans to buy or to build a home | | | **x** |
| Likelihood to spend large sums of money on home improvements or renovations | | | **x** |

3.41 Additional questions could be considered. For instance, consumer surveys in the US, carried out by the Thomson-Reuters-University of Michigan and by the Conference Board, also include questions about the subjective probability assigned by consumers to the realisation of specific events (income growth being higher or lower than inflation, falling into unemployment, adequateness of the retirement scheme) and questions on interest rate expectations and price expectations for specific goods (gasoline and house prices).

# Annex

## Questionnaires with complete questions

## Table 1

## Monthly BTS

**INDUSTRY**

**Evolution of production / business activity - past 3 months**
How has your production/ business activity changed over the past 3 months?
[It has (1) increased, (2) remained unchanged, (3) decreased]

**Evolution of production / business activity - next 3 months**
How do you expect your production/ business activity to change over the next 3 months?
[It will (1) increase, (2) remain unchanged, (3) decrease]

**Evolution of employment - past 3 months**
How has your firm's total employment changed over the past 3 months?
[It has (1) increased, (2) remained unchanged, (3) decreased]

**Evolution of employment - next 3 months**
How do you expect your firm's total employment to change over the next 3 months?
[It will (1) increase, (2) remain unchanged, (3) decrease]

**Level of order books - past 3 months**
3 months ago, did you consider your overall order books to be...?
[(1) Above normal, (2) normal, (3) below normal]

**Level of order books - current**
Do you consider your current overall order books to be..?
[(1) Above normal, (2) normal, (3) below normal]

**Level of export order books - past 3 months**
3 months ago, did you consider your level of export order books to be...?
[(1) Above normal, (2) normal, (3) below normal]

**Level of export order books - current**
Do you consider your current level of export order books to be..?
[(1) Above normal, (2) normal, (3) below normal]

**Stock of finished goods - current**
Do you consider your current stock of finished products to be..?
[(1) Above normal, (2) normal, (3) below normal]

**Evolution of selling prices - next 3 months**
How do you expect your selling prices to change over the next 3 months?
[It will (1) increase, (2) remain unchanged, (3) decrease]

**CONSTRUCTION**

**Evolution of business activity - last 3 months**
How has your business activity changed over the past 3 months?
[It has (1) increased, (2) remained unchanged, (3) decreased]

**Evolution of business activity - next 3 months**
How do you expect your business activity to change over the next 3 months?
[It will (1) increase, (2) remain unchanged, (3) decrease]

**Evolution of employment - last 3 months**
How has your firm's total employment changed over the past 3 months?
[It has (1) increased, (2) remained unchanged, (3) decreased]

**Evolution of employment - next 3 months**
How do you expect your firm's total employment to change over the next 3 months?
[It will (1) increase, (2) remain unchanged, (3) decrease]

**Level of order books - past 3 months**
3 months ago, did you consider your overall order books to be...?
[(1) Above normal, (2) normal, (3) below normal]

**Level of order books - current**
Do you consider your current overall order books to be..?
[(1) Above normal, (2) normal, (3) below normal]

**Factors limiting production - current**

What main factors are currently limiting your building activity?
[ (1) None, (2) insufficient demand, (3) weather conditions, (4) shortage of Labour Force, (5) shortage of material, (6) financial constraints, (7) other factors]

**Evolution of selling prices - next 3 months**
How do you expect your selling prices to change over the next 3 months?
[They will (1) increase, (2) remain unchanged, (3) decrease]

## RETAIL TRADE

**Evolution of sales / business activity - last 3 months**
How has your business activity (sales) changed over the past 3 months?
[It has (1) increased, (2) remained unchanged, (3) decreased]

**Evolution of sales / business activity - next 3 months**
How do you expect your business activity (sales) to change over the next 3 months?
[It will (1) increase, (2) remain unchanged, (3) decrease]

**Evolution of employment - last 3 months**
How has your firm's total employment changed over the past 3 months?
[It has (1) increased, (2) remained unchanged, (3) decreased]

**Evolution of employment - next 3 months**
How do you expect your firm's total employment to change over the next 3 months?
[It will (1) increase, (2) remain unchanged, (3) decrease]

**Evolution of orders placed with suppliers - next 3 months**
How do you expect your orders placed with suppliers to change over the next 3 months?
[They will [(1) increase, (2) remain unchanged, (3) decrease]

**Stock of finished goods - current**
Do you consider your current stock of finished products to be..?
[(1) Above normal, (2) normal, (3) below normal]

**Evolution of selling prices - next 3 months**
How do you expect your selling prices to change over the next 3 months?
[They will (1) increase, (2) remain unchanged, (3) decrease]

## FINANCIAL SERVICES

**Evolution of business activity - last 3 months**
How has your business activity changed over the past 3 months?
[It has (1) increased, (2) remained unchanged, (3) decreased]

**Evolution of business activity - next 3 months**
How do you expect your business activity to change over the next 3 months?
[It will (1) increase, (2) remain unchanged, (3) decrease]

**Evolution of employment - last 3 months**
How has your firm's total employment changed over the past 3 months?
[It has (1) increased, (2) remained unchanged, (3) decreased]

**Evolution of employment - next 3 months**
How do you expect your firm's total employment to change over the next 3 months?
[It will (1) increase, (2) remain unchanged, (3) decrease]

**Evolution of demand / turnover - last 3 months**
How has demand for your company's services changed over the past 3 months?
[It has (1) increased, (2) remained unchanged, (3) decreased]

**Evolution of demand / turnover - next 3 months**
How do you expect the demand for your company's services to change over the next 3 months?
[It will (1) increase, (2) remain unchanged, (3) decrease]

## OTHER SERVICES

**Evolution of business activity - last 3 months**
How has your business activity changed over the past 3 months?
[It has (1) increased, (2) remained unchanged, (3) decreased]

**Evolution of business activity - next 3 months**
How do you expect your business activity to change over the next 3 months?
[It will (1) increase, (2) remain unchanged, (3) decrease]

**Evolution of employment - last 3 months**
How has your firm's total employment changed over the past 3 months?
[It has (1) increased, (2) remained unchanged, (3) decreased]

**Evolution of employment - next 3 months**
How do you expect your firm's total employment to change over the next 3 months?
[It will (1) increase, (2) remain unchanged, (3) decrease]

**Evolution of demand / turnover - last 3 months**
How has demand for your company's services changed over the past 3 months?
[It has (1) increased, (2) remained unchanged, (3) decreased]
**Evolution of demand / turnover - next 3 months**
How do you expect the demand for your company's services to change over the next 3 months?
[It will (1) increase, (2) remain unchanged, (3) decrease]
**Evolution of selling prices - next 3 months**
How do you expect your selling prices to change over the next 3 months?
[They will (1) increase, (2) remain unchanged, (3) decrease]

**WHOLESALE**
**Evolution of sales / business activity - last 3 months**
How has your business activity (sales) changed over the past 3 months?
[It has (1) increased, (2) remained unchanged, (3) decreased]
**Evolution of sales / business activity - next 3 months**
How do you expect your business activity (sales) to change over the next 3 months?
[It will (1) increase, (2) remain unchanged, (3) decrease]
**Evolution of employment - last 3 months**
How has your firm's total employment changed over the past 3 months?
[It has (1) increased, (2) remained unchanged, (3) decreased]
**Evolution of employment - next 3 months**
How do you expect your firm's total employment to change over the next 3 months?
[It will (1) increase, (2) remain unchange, (3) decrease]
**Stock of finished goods - current**
Do you consider your current stock of finished products to be..?
[(1) Above normal, (2) normal, (3) below normal]
**Evolution of selling prices - next 3 months**
How do you expect your selling prices to change over the next 3 months?
[They will (1) increase, (2) remain unchanged, (3) decrease]

**AGRICULTURE**
**Evolution of production / business activity - last 3 months**
How has your business activity changed over the past 3 months?
[It has (1) increased, (2) remained unchanged, (3) decreased]
**Evolution of production / business activity - next 3 months**
How do you expect your production/business activity to change over the next 3 months?
[It will (1) increase, (2) remain unchanged, (3) decrease]
**Evolution of employment - last 3 months**
How has your firm's total employment changed over the past 3 months?
[It has (1) increased, (2) remained unchanged, (3) decreased]
**Evolution of employment - next 3 months**
How do you expect your firm's total employment to change over the next 3 months?
[It will (1) increase, (2) remain unchanged, (3) decrease]
**Factors limiting production – current**
What main factors are currently limiting your agriculture activity?
[(1) None, (2) insufficient demand, (3) weather conditions, (4) shortage of Labour Force, (5) shortage of material, (6) financial constraints, (7) other factors]
**Stock of finished goods – current**
Do you consider your current stock of finished products to be..?
[(1) Above normal, (2) normal, (3) below normal]
**Evolution of selling prices - next 3 months**
How do you expect your selling prices to change over the next 3 months?
[They will (1) increase, (2) remain unchanged, (3) decrease]

## Table 2

## Quarterly BTS

**INDUSTRY**
**Evolution of order books - past 3 months**

How have your orders changed over the past 3 months?
[They have (1) increased, (2) remained unchanged, (3) decreased]

**Evolution of order books - next 3 months**
How do you expect your order books to change over the next 3 months?
[They will (1) increase, (2) remain unchanged, (3) decrease]

**Evolution of export order books - next 3 months**
How do you expect your export orders to change over the next 3 months?
[They will (1) increase, (2) remain unchanged, (3) decrease]

**Months of production assured by current order books**
How many months of production are assured by your current order books?
[please provide number of months]

**Factors limiting production - current**
What main factors are currently limiting your production?
[(1) None, (2) insufficient demand, (3) shortage of labour force, (4) shortage of material, (5) financial constraints, (6) other factors]

**Production capacity - current**
Considering your current order books and the expected change in demand over the coming months, how do you assess your current production capacity?"
[(1) more than sufficient, (2) sufficient, (3) not sufficient]

**Capacity utilisation - current**
At what capacity is your company currently operating (as a % of full capacity)?
[please provide %]

**Stock of raw materials - current**
Do you consider your current stock of raw materials to be..?
[(1) Above normal, (2) normal, (3) below normal]

**Evolution of the competitive position on the domestic market - past 3 months**
How has your competitive position changed over the last 3 months?
[It has (1) increased, (2) remained unchanged, (3) decreased]

## CONSTRUCTION

**Months of production assured by current order books - current**
How many months of production are assured by your current order books?
[please provide number of months]

**Capacity utilisation - current**
At what capacity is your company currently operating (as a % of full capacity)?
[please provide %]

## FINANCIAL SERVICES

**Evolution of operating income - past 3 months**
How has your operating income changed over the last 3 months?
[It has (1) increased, (2) remained unchanged, (3) decreased]

**Evolution of operating income - next 3 months**
How do you expect your operating income to change over the next 3 months?
[It will (1) increase, (2) remain unchanged, (3) decrease]

**Evolution of operating expenses - past 3 months**
How have your operating expenses changed over the last 3 months?
[They have (1) increased, (2) remained unchanged, (3) decreased]

**Evolution of operating expenses - next 3 months**
How do you expect your operating expenses to change over the next 3 months?
[They will (1) increase, (2) remain unchanged, (3) decrease]

**Evolution of profitability - past 3 months**
How has your profitability changed over the last 3 months?
[It has (1) increased, (2) remained unchanged, (3) decreased]

**Evolution of profitability - next 3 months**
How do you expect your profitability to change over the next 3 months?
[It will (1) increase, (2) remain unchanged, (3) decrease]

**Evolution of capital expenditure - past 3 months**
How has your capital expenditure changed over the last 3 months?
[It has (1) increased, (2) remained unchanged, (3) decreased]

**Evolution of capital expenditure - next 3 months**

How do you expect your capital expenditure to change over the next 3 months?

[It will (1) increase, (2) remain unchanged, (3) decrease]

**Evolution of competitive position - past 3 months**
How has your competitive position changed over the last 3 months?
[It has (1) increased, (2) remained unchanged, (3) decreased]

**Evolution of competitive position - next 3 months**
How do you expect your competitive position to change over the next 3 months?
[It will (1) increase, (2) remain unchanged, (3) decrease]

**OTHER SERVICES**

**Factors limiting activity - current**
What main factors are currently limiting your activity?
[(1) None, (2) insufficient demand, (3) shortage of labour force, (4) shortage of space/ equipment, (5) financial constraints, (6) other factors]

## Table 3

## (Semi-)Annual BTS

**INDUSTRY**

**Development of fixed investment - last year**
Please state percentage change in investment last year (t-1) on investment 2 years ago (t-2): [%]

**Development of fixed investment - this year**
Please state percentage change in investment this year (t) on investment last year (t-1): [%]

**Development of fixed investment - next year**
Please state percentage change in investment next year (t+1) on investment this year (t): [%]

**CONSTRUCTION**

**Development of fixed investment - last year**
Please state percentage change in investment last year (t-1) on investment 2 years ago (t-2): [%]

**Development of fixed investment - this year**
Please state percentage change in investment this year (t) on investment last year (t-1): [%]

**Development of fixed investment - next year**
Please state percentage change in investment next year (t+1) on investment this year (t): [%]

**WHOLESALE**

**Development of fixed investment - last year**
Please state percentage change in investment last year (t-1) on investment 2 years ago (t-2): [%]

**Development of fixed investment - this year**
Please state percentage change in investment this year (t) on investment last year (t-1): [%]

**Development of fixed investment - next year**
Please state percentage change in investment next year (t+1) on investment this year (t): [%]

**AGRICULTURE**

**Development of fixed investment - last year**
Please state percentage change in investment last year (t-1) on investment 2 years ago (t-2): [%]

**Development of fixed investment - this year**
Please state percentage change in investment this year (t) on investment last year (t-1): [%]

**Development of fixed investment - next year**
Please state percentage change in investment next year (t+1) on investment this year (t): [%]

## Table 4

## Monthly CTS

**Evolution of household financial situation - last 12 months**
How has the financial situation of your household changed over the last 12 months?
[It has… (1) got a lot better, (2) got a little better, (3) stayed the same, (4) got a little worse, (5) got a lot worse, (6) don't know]

**Evolution of household financial situation - next 12 months**
How do you expect the financial situation of your household to change over the next 12 months?
[It will...(1) get a lot better, (2) get a little better, (3) stay the same, (4) get a little worse, (5) get a lot worse, (6) don't know]

**Evolution of the general economic situation in the country - last 12 months**
How do you think the general economic situation in the country has changed over the past 12 months?
[It has...(1) got a lot better, (2) got a little better, (3) stayed the same, (4) got a little worse, (5) got a lot worse, (6) don't know]

**Evolution of the general economic situation in the country - next 12 months**
How do you expect the general economic situation in the country to change over the next 12 months?
[It will...(1) get a lot better, (2) get a little better, (3) stay the same, (4) get a little worse, (5) get a lot worse, (6) don't know]

**Evolution of consumer prices - last 12 months**
How do you think that consumer prices have changed over the last 12 months?
[They have... (1) risen a lot, (2) risen moderately, (3) risen slightly, (4) stayed the same, (5) fallen, (6) don't know]

**Evolution of consumer prices - next 12 months**
How do you expect the consumer prices to change over the next 12 months?
[They will... (1)increase more rapidly, (2) increase at the same rate, (3) increase at a slower rate, (4) stay about the same, (5) fall, (6) don't know]

**Evolution of the number of unemployed people - next 12 months**
How do you expect the number of unemployed in this country to change over the next 12 months?
[They will...(1) increase more sharply, (2) increase slightly, (3) remain the same, (4) fall slightly, (5) fall sharply, (6) don't know]

**Right moment to make major purchases - current**
Considering the general economic situation, do you think that now it is the right moment for people to make major purchases such as furniture, electrical/electronic devices, etc?
[(1) yes it is, (2) it's neither the right nor the wrong moment, (3) no it is not, (4) don't know]

**Evolution of money spent on major purchases - next 12 months**
Compared to the past 12 months, do you expect to spend more or less money on major purchases (furniture, electrical/electronic devices, etc) over the next 12 months?
[I will spend… (1) much more, (2) a little more, (3) about the same, (4) a little less, (5) much less, (6) don't know]

**Opportunity to save - current**
Considering the general economic situation, do you think that now it is..?
[(1) a very good moment to save, (2) a fairly good moment to save, (3) not a good moment to save, (4) a very bad moment to save, (5) don't know]

**Opportunity to save - next 12 months**
Over the next 12 months, how likely it is that you save any money?
[(1) very likely, (2) fairly likely, (3) not likely, (4) not at all likely, (5) don't know]

**Statement best describing the household financial situation - current**
Which of these statements best describes the current financial situation of your household?
[We are…(1) saving a lot, (2) saving a little, (3) just managing to make ends meet on our income, (4) having to draw on our savings, (5) running into debt, (6) don't know]

## Table 5

## Quarterly CTS

**Likelihood to buy a car - next 12 months**
How likely would you buy a car in the next 12 months?
[(1) very likely, (2) fairly likely, (3) not likely, (4) not at all likely, (5) don't know]

**Plans to buy or to build a home - next 12 months**

Are you planning to buy or build a home in the next 12 months (to live in yourself, a member of your family, as a holiday home, to let, etc.)?
[(1) yes definitely, (2) possibly, (3) probably not, (4) no), (5) don't know]

**Likelihood to spend large sums of money on home improvements or renovations - next 12 months**

How likely would you spend any large sums on home improvements or renovations over the next 12 months?
[(1) very likely, (2) fairly likely, (3) not likely, (4) not at all likely, (5) don't know]

## References

OECD (2003). *Business Tendency Surveys: A handbook.* OECD Publications, Paris, ISBN 92-64-19894-6. (http://www.oecd.org/std/leading-indicators/31837055.pdf.

OECD study (2010*). Is it necessary to seasonally adjust business and consumer confidence series?* (http://www.oecd.org/std/leading-indicators/43858762.pdf), OECD, Paris.

European Commission, Directorate-General for Economic and Financial Affairs (2014), *"The Joint Harmonised EU Programme of Business and Consumer Surveys"*, user guide March, available at http://ec.europa.eu/economy_finance/db_indicators/surveys/method_guides/index_en.htm.

# CHAPTER 4

## Survey frame and sample design

### A. Introduction

4.1 The previous chapters have dealt with the general scope and use of ETS and the desired output. In this chapter and the next ones we will discuss the input that is needed to produce this output, and how the input must be processed to arrive at the output. More specifically, in this chapter we focus on important steps in the entire process of ETS: from where and how to select a subset of units (i.e. the sample) upon which we can draw estimates of the information we wish to collect (i.e. variables composing confidence).

Based on the limited resources available, normally an organisation cannot do a comprehensive survey of all businesses/consumers. Instead, one must conduct a sample survey, which naturally leads to sample design, the main topic of this chapter.

First of all, it is worth mentioning that in setting up a sample survey, there is no unique or optimal solution. Wide choices of sampling methods are available and all could be worth depending on the starting hypotheses set, the targets to reach and the limits to face. Furthermore other steps of the entire process, like selection of the survey frame or the weighting system to adopt, affects (or are affected by) the sample.

Before dealing with sampling designs the proper identification and selection of other statistical entities is required: namely the sampling unit, the target population and the survey frame, which is basically a list of all units of the target population. In the following we will first discuss some of the issues regarding the survey frame and its constituent units.

### B. Survey frame

4.2 In setting up a survey the first problem we have to solve is to clearly identify the population from which we wish to collect information: this is generally defined as target universe and may comprise all or only the subset of interest of the overall population. In case of a business survey, e.g., we can focus our target universe on the enterprises belonging to some NACE codes, like manufacturing or services; for a consumer survey, e.g. , we could wish to contact individuals belonging only to some classes of age (e.g. full aged persons).

We also have to decide which units are going to be contacted, that is the physical units (for business: firms, establishments, etc.; for consumers: voters, telephone subscribers, etc.) we can materially contact: the reporting units. The sample unit is more properly the unit on which we try to collect the statistical information. Reporting and sample units may coincide or not. For example, an enterprise is contacted (reporting unit) to collect statistical information on its local units (sampling units), or, for consumer surveys, household is contacted (reporting unit) to collect information on consumers (sampling unit).

The selection of most suitable register, the frame list, is then affected by the previous identification of the above mentioned entities and should comprise all and only those sampling units belonging to the target population. Furthermore, as prerequisite for conducting a survey, a reliable, complete and up-to-date list of all possible subjects of investigation, as far as possible, has to be available, so that all units have a non -zero chance being selected.

To ensure a high quality of the survey, the frame list should then be frequently updated in order to monitor as close as possible the evolution of the relevant population and ensure a good coverage, minimising missing and duplicate entries.

**(i) Business tendency surveys**

4.3 The ideal coverage of BTS for each sector requires to include all enterprises engaged in the relevant kind of activity, regardless of their type of ownership, legal form, size, and whether or not one of the kinds of activity is their predominant or secondary activity. In addition the target population should cover all enterprises active at any time during the period covered in that survey, including any forecast period[13].

4.4 As the selection of the most suitable register is affected by the decision on which sample units to use, we first describe the different types of units that can be distinguished. At the most fundamental level we find the establishment unit. An establishment is an enterprise, or part of an enterprise, that is situated in a single location and in which only a single (non-ancillary) productive activity is carried out or in which the principal productive activity accounts for most of the value added.

The establishment units contained in a single enterprise can be grouped together in different ways:

- All establishments at the same location together constitute the local unit
- All establishments carrying out the same industrial activity together constitute the kind-of-activity unit (KAU)

[13] See chapter 7

- All establishments together constitute the enterprise unit. This unit comprises "all activity in all locations carried out by one legal entity.

Many companies, especially the smaller ones, are located in just one location and their whole activity is in one kind of activity. For these the establishment, local unit, kind of activity unit and enterprise are the same.

4.5 The choice of the sample unit can influence survey results as follows:

- If the survey uses establishments as sampling units, the aggregates by industry will cover all activity in each industry. In addition, aggregates for different regions will contain all the activity for that industry in that region and no activity in other regions. In this case the sample unit coincides with the reporting one. Basing a survey on establishment units is the ideal situation
- If local units are used as the sampling unit, they have to be classified by their main activity. Therefore, aggregates by industry will contain both too much and too little at the same time. They will include activity in industries other than the main activity of the local units, while at the same time excluding activity in that industry at local units mainly engaged in other industries. Aggregation by region will give correct information, i.e. the same result as if establishments had been used as the reporting unit
- If kind–of–activity units are used as the sampling units, the aggregates by industry will be the same as if establishments had been used. On the other hand, aggregates by region cannot be derived without simplifying assumptions

If the whole enterprise unit is used as the sampling unit, aggregates by industry will include activity in industries other than the main one, and exclude activity in that industry of enterprises mainly engaged in other industries. This problem with under–coverage and over-coverage will often be more pronounced if the enterprise is used as the reporting unit than when local units are used. The reason for this is that one can expect a larger and more pronounced diversity of industrial activity within an enterprise than in a local unit. To decide which unit to use one should refer to the output specification. The output specification should describe if regional and / or kind of activity output data are required. From the above overview one can then derive which units should ideally be used.
It should be kept in mind, though, that multi-establishment enterprises are often unable to provide reliable information at the establishment level. Therefore it may be impossible to use the establishment units, even though these would be the ideal choice.

The survey frame is usually composed by business registers which record the addresses and contact details of enterprises, whether corporate or unincorporated, list their main kind of economic activity and give some measure of their size – usually the number of employees. Business registers may also contain information on the different establishments owned by each enterprise, which may be located in different parts of the country and which may be involved in different kinds of activity. Establishing a business register and keeping it up-to-date by adding new enterprises and eliminating those that have ceased operations, is a major task in most statistical offices. A good summary of best practices in designing and maintaining business registers is given in the handbook of OECD(2003). More detailed guidance is available in Eurostat publications (2003) about business register recommendations. Here we assume that there is a reasonably comprehensive business register and that, at a minimum, it contains information on the addresses of enterprises, their main kinds of activity and their approximate size.

Many BTS are carried out by chambers of commerce or employers' associations and use their membership list as the survey frame. Note that such frames are affected by self-selection bias: such surveys can only provide information about the enterprises that belong to the association and their use for monitoring developments in the entire industrial sector depends on the extent to which the member enterprises are representative of the sector as a whole. In practice, many surveys based on membership lists have been shown to provide reliable information on movements in total output or industrial production, but the recommendation of this Handbook is that BTS should use as their survey frame comprehensive business registers of the kind that national statistical agencies maintain for their regular enterprise surveys, if available.

**(ii) Consumer tendency surveys**

4.7 In CTS the target population usually represents only a subset of the entire population. Cut-off on the basis of age is usually applied in EU, the cut-off age varies across countries (with a minimum age of 14-18 years, to investigate only full aged individuals, and a maximum age in some cases is applied too); in some countries, cut-offs on geographical basis are also applied and in one case also for working status.

As confidence is indirectly monitored by asking questions regarding present and expected situation of the consumer himself and of his household, some uncertainties in defining the reference population may arise, that is: do we have to consider as sample unit the consumer or the household

to which he belongs, or both? This question has consequences on the frame to use, on the kind of sample to select, and on the estimation of the results.
In the consumer survey of the European Harmonised Programme the sample unit is represented by the consumer. The subject of investigation, that is the sample unit, is represented by the consumer intended as a full aged person contributing also in non-monetary terms to the household management. In some countries however, the sample unit is represented by the household. Some doubt can however arise when contacting a member of an household while we are asking opinions (as in EU harmonised consumer survey) both on personal and on household situation. The response unit, that is the individual who is materially interviewed, usually does not differ from the sample unit, when it is defined as consumer, while when dealing with households we face the task to select a person within the family unit.
It is worth to remember that, up to 1994 the harmonised consumer survey was focused on households and currently some questions are still referred to household. This occurrence is to be kept in mind while handling this survey.

4.8 In consumer surveys, frame lists can be composed by individuals like population register, or electoral rolls, but also referring to individuals like mobile phone register. On the other side, lists can deal with dwellings/households, like fixed phone registers. The choice of the list, apart from the subject of investigation, can be related to several other motivations, like effective availability and to the data collection mode selected. If, e.g., the survey is conducted via a Computer Aided Telephone Interviewing (CATI) system, a phone register will be the suitable frame list, if the survey is conducted face-to-face a population register will be the most advisable frame list.

4.9 Fixed telephone lists still represent the most used ones. They are however affected by some limits: the fixed phone directories refer to dwellings, that is households, not to individuals; this occurrence affects the sampling design. Furthermore, in these lists increasing risks of under-coverage are present: possible exclusion of part of the population who cannot afford the phone cost, or because of privacy reasons, as in some countries individuals may choose not to be comprised in the telephone directories (so called “opposition list” or “red list”). But today the most important reason of under-coverage is due to the growing switching to mobile phones.
Mobile phones have also some limits, as pointed out by R. Curtin (2003, 2010). Generally, an interview conducted via mobile phone is more difficult as respondent can be busy and not able to answer, also with risks for their safety (e.g. contacted while driving). Mobile phones are less used by elderly people for digital divide reasons. As the possibility to maintain the same number while

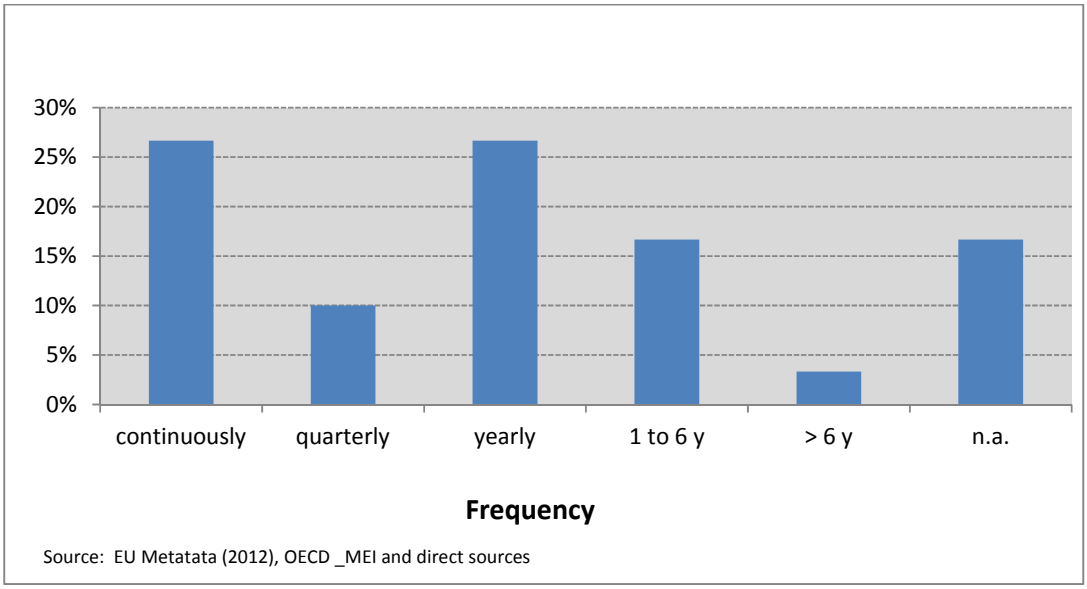
changing company - and in some countries also when switching from fix to mobile - is increasing, it is not anymore possible to recognize the geographical location of the phone number preventing from possible stratification of the frame lists.

The use of mobile phones could have an unfavourable impact on respondent selection since most cellular phones are associated with specific household members. The equal probability methods, advisable for selecting households, must be followed by the selection of a specific person as respondent also by using a probabilistic selection rule. This is easy to implement when all household members have access to the same landline phone, but it is quite difficult with mobile phones.

Even though internet surveys are strongly increasing, lists based on internet connections presently show several drawbacks, most of all the lack of coverage of the entire population, the inability to select a representative sample of households, and the inability to verify which person within the household actually answered the questions. (These topics are dealt in more details in chapter 6) .

**Figure F4.1**

**Frame list update**




According to the 2012 information, in OECD countries, about 2/3 of the institutes update the consumers frame lists at least yearly (of which 27% continuously and this quota is even increasing in recent years), in almost 20% of cases the lists are however updated every two or even 5-6 years and 17% do not declare the frequency (see Figure F4.1).

## C. Sample design

4.10 A sample is a subset of the frame. Instead of surveying all units in the frame, only those in the sample are surveyed. The output is produced using the data from the surveyed units. Because not all units are surveyed, the output consists of estimates of the "real", unknown population value. A question that arises naturally from this is how precise these estimates are. Only in the case of probabilistic samples (see following subparagraph) the precision of these estimates can be determined.

The sample design comprise various features so as to increase efficiency and to accomplish the research aims, mainly consisting in the mode of units' selection, the setup of the sample structure and the allocation criteria of the selected unit to the chosen structure of the sample. Hereafter a concise description of these aspects follows; a thorough treatment of sampling theory, with its multitude of different sample designs, is beyond the scope of this Handbook. Readers who want to know more about this fascinating topic are referred to the literature.

### (i) Selecting units: probabilistic and deterministic samples

4.11 A *probabilistic sample* selection criterion means that the units to include in the sample are randomly selected from the frame using a probabilistic algorithm. Every element of the frame has a known positive chance to be included so that it is possible to compute the probability of every possible distinct sample selected. It allows normality assumptions and consequently to calculate the precision of estimates (see chapter 5). The random criterion for selecting sample units is the principle on which the majority of ETS samples relies.

A further feature, the Equal Probability of Selection Method (EPSEM), characterizes probabilistic samples whose units are selected with a probability criterion so that all the units of the population have equal probabilities of being included in the sample. EPSEM samples are self-weighting that is, the weights don't need to be calculated separately (see chapter 5).

In subparagraph C(ii) a concise but exhaustive discussion of the most applied samples designs based on probabilistic selection, best suitable for ETS, is presented.

4.12 For sake of completeness, a short introduction to deterministic samples is also hereafter presented. The deterministic or non-probability selection techniques, select units in different ways with respect to random selection. In *deterministic samples*, $n$ units are selected out of the $N$ from which the population is composed using any deterministic (non-random) selection criterion. The inference from the sample to population is therefore generally not possible. This in turn implies that when using a deterministic sample, it is impossible to ascertain that the output satisfies the specified

precision. Furthermore, even samples initially built as probability samples sometimes end up being a non-probability one due to unintentional or unavoidable addition to the selected sampling design. The most applied samples based on deterministic techniques of selection units, mainly applied for consumer surveys, are:

- *purposive* or *judgmental samples*: units are selected based on knowledge of a population and on the purposes of the study. Units are selected according to some characteristic. These kinds of samples are selected for studying particular subsets of population or extreme cases (e.g. for individuals: illnesses, social diseases, etc.). Generally, purposive samples are not suitable for ETS

- *the quota sample* represents the most used of the non-probabilistic methods, mainly applied in opinion polls and market researches. Quotas are built by subdividing the population in homogeneous classes according to some variables (gender, age, address, etc.). The total sample size is thus spread within the classes so as to mirror the corresponding population weights. Quotas are then defined as the number of interviews to be carried out in each class. The main feature of this kind of sampling is that the units to be interviewed are directly selected by the interviewer in the frame of the quotas assigned. Quota sampling assumes that those who take part in the survey have the same characteristics, attitudes, behaviours etc. as those who do not take part and this occurrence is all but not true. The quota definition can be made at different steps of the sampling plan. For example gender quotas can be set within random sample of consumer. In these cases it deals more properly of probability sample with quotas

- a *substitution sample* is a quota sample in which the units are selected following probability techniques. Practical experience has shown that, even cautiously, the usual statistical methods can still be applied

- finally, it is worth mentioning *Random Digit Dialing (RDD) techniques*: a set of techniques for selecting sample units by generating telephone numbers at random. RDD has the advantage that it includes unlisted numbers that would be missed if the numbers were selected from a phone register. In populations where there is a high telephone-ownership rate, it can be a cost efficient way to get complete coverage of a geographic area. The method is widely applied in the USA and the specific procedure applied by the Survey Research Center of the Michigan University is reported in Curtin, 2013.

**(ii) Sample design features**

4.13 Many different sample designs exist as they depend on the particular circumstances for which the design is the most efficient. Some oft-used sample designs based on probabilistic unit selection with their most important properties are described hereinafter.

*Simple Random Sampling* – SRS (strictly speaking Simple Random Sampling Without Replacement) is one of the most straightforward and familiar probabilistic sample designs. In a SRS Design, all possible samples of equal size from the survey frame have exactly the same chance of selection. Conceptually, SRS can be implemented by writing the name of each unit in the survey frame on a separate piece of paper, putting all these pieces of paper in a box and drawing the desired number of pieces of paper from the box at random. For obvious reasons, in practice one uses a computer, but the idea is the same: make a list of all units in the frame, draw one of them at random, add it to the sample and remove it from the list, and repeat until the sample has the desired size. SRS isn't often used on its own as a sample design in ETS. But all other random sampling techniques can be viewed as an extension or adaptation of this method, so it is worth noting nonetheless. Furthermore, this SRS is often compared with more complex designs (see further) as benchmark to test their efficiency.

In *Stratified Random Sampling* one first divides the frame into several non-overlapping sub-frames called strata, and uses a SRS design in each stratum. Although conceptually simple, in practice, arriving at an efficient Stratified Sampling design can be quite complex: one not only needs to consider carefully which strata to use, but also how many units to include in the sample within each stratum (this question is also known as the allocation problem, see hereinafter).

Stratified Random Sampling can be used for a variety of reasons:

1. stratification is a common way of improving the precision of estimators. If it is possible to form strata within which the target variable varies little, stratified sampling can lead to more precise outcomes than SRS (if the sample size is the same). It can also be a more efficient design: if the desired precision is the same, in a stratified sampling design the required sample size is often smaller than for SRS
2. the interest is often not only in the population as a whole, but also in specific subpopulations or in making comparisons between subpopulations. In SRS, it is a matter of chance how many elements end up in the strata. Small subpopulations in particular will then be poorly represented in the sample. Stratification is a way of ensuring that all subpopulations of interest are sufficiently represented in the sample to allow reliable statements to be made
3. it is possible in stratification to use different data collection techniques for different strata. For instance, in BTS it may be desirable to approach small companies by means of a brief

paper questionnaire and to have large companies take part in a telephone or personal interview. The selection and estimating methods may also differ for each stratum

4. for administrative reasons, sampling frames are often already divided into 'natural' parts, which may even be kept at geographically different locations. In this case separate sampling may be more economical.

4.14 If the output specification requires a minimum level of precision for the estimators, we can determine the necessary stratum sample sizes using the methods we will discuss in the next chapter. Collectively these sub samples form the total sample. Often, however, the total sample size will be fixed, and the question is how to distribute these elements over the strata arise. This is known as the allocation problem. Different allocation methods exist in the literature, the two more applied are presented.

4.15 In *proportional allocation*, the size of the sample in each stratum is taken in proportion to the size of the stratum. If, e.g., the overall sample size is 1 per cent of the survey frame size, the stratum sample size is 1 per cent of the total number of units in each stratum. All elements in the frame, irrespective of stratum, have the same probability of being selected in the sample. As a consequence of this, as we will see in chapter 5, all units have the same weight in the estimation. We refer in this case to a self-weighting sample.

We note that in this allocation scheme, the inclusion probability (the chance of a given unit being drawn in the sample) is the same as for SRS. Therefore, it may come as a surprise that proportional allocation often gives more precise estimators. The reason for this is that the sampling error depends on the variance within each stratum, and this variance will be lower than the overall population variance when the strata are sufficiently homogeneous.

When the variances within various stratum are equal, proportional allocation is the best allocation method for improving precision. This, however, is not often the case: larger businesses will generally exhibit greater variability on the target variable than the smaller ones. An allocation method that takes these differences in variability into account is therefore often more efficient in practice; one such method is called Neyman allocation.

4.16 If the stratum variances vary strongly, *optimum allocation to strata* (also called *Neyman allocation*) of the sample leads to greatest precision of the estimators. In this allocation method the sample size in each stratum is proportional not only to the number of units in the stratum, but also to the in-stratum variance. Note that this can lead to a stratum size larger than the corresponding frame stratum size: this simply means that the entire stratum has to be included in the sample.

Determining the optimum allocation requires that all stratum variances are known. This information will seldom be available in practice, but the variances can sometimes be approximated based on earlier surveys. When no data from earlier surveys is available, one could first perform a survey based on a simple sample design (e.g., SRS or proportional allocation), and successively optimize the design in later periods by using the data gathered from the earlier survey.

4.17 A further design, relevant to BTS, in contrast with the sample designs discussed so far, is the *Probability Proportional to Size* (PPS)sample design, where not all units have the same inclusion probability. Instead, as the name suggests, the probability of inclusion of a unit in the sample is proportional to the size of the unit. Strictly speaking, any auxiliary variable for which the value is known for each unit in the frame can be used as the size, but in practice the number of employees is the most obvious and the most popular choice.

The advantage of using a PPS design is that the resulting estimators can be much more precise than for a simple random sample of the same size. Especially in cases where there is a close relation between the target variable and the size variable, PPS designs can be very efficient. Another advantage is that stratification in size classes is not necessary. This lessens the risk of having extremely few observations in some strata when the total sample size is fixed (see the "rule of 30").

Conceptually, a PPS design can be implemented by first cutting out a strip of paper for each unit of the frame, where the length of each strip is proportional to the size of the unit, and writing the name of the unit on it. Then lay these strips of paper end-to-end on the floor in a random order, in such a way that they form one long ribbon. Next, determine a parameter called the step length by dividing the total length of the ribbon by the required sample size, and generate a random number between 0 and the step length. This random number determines the starting point: measure out a distance equal to this random number from the left-most edge of the ribbon and mark it on the ribbon. Then repeatedly make additional marks on the ribbon, each one a step length to the right of the previous mark, until you run out of ribbon. Finally, break up the ribbon into individual strips of paper. The sample now consists of all units on strips of paper with a mark on them.

Note that this procedure only results in the correct number of units if no strips of paper are wider than the step length: A strip of paper wider than the step length would receive multiple marks. It is tacitly assumed that all units larger than the step length have been removed from the frame and included in a separate stratum, which is observed in full. These units are referred to as the self-selecting units. The above procedure is then performed for the remaining ones. Because the step length is recalculated after removing self-selecting units, additional self-selecting units may appear. These must be repeatedly removed until no new self-selecting units appear.

4.18 Obviously, PPS sampling can be combined with the idea of stratification. For example, the actual procedure used by Statistics Netherlands is to first stratify the survey frame according to kind of activity, and then use a PPS sample within each stratum.

4.19 Other features that can affect samples designs used for ETS, in particular for consumer survey, are:

a) *Stage sample design*: introducing stages means that the final units, e.g. consumers, whenever selected (randomly or not), are not immediately chosen from population, but trough subsequent steps, e.g.: for CTS at the first stage household are selected (primary units) , at the second the consumers (secondary units).

b) *Cluster sampling* consists in selecting samples composed of subsets of the reference population, featured as small scale representation of the total population. The single-stage cluster usually imply the sampling of all units comprised within the selected subset as the sample unit is represented by the cluster itself. In two(/multi)-stage cluster sampling, instead, a random sampling technique is applied to collect the desired units from each of the selected clusters. Clustering is adopted in survey sampling for practical reasons. Mainly, if we have to build the sample into manageable workloads for interviewers, while using the face-to-face data collecting mode. This feature, while allows saving of travelling time and consequent reduction in costs, however implies larger sampling error than un-clustered samples, as only a part of the population is monitored.

4.20 Less widespread samples designs based on probabilistic selection techniques are:

a) *Balanced sampling*: is a random method of selection of units from a population that is applied to stratified designs and to fixed size samples so that the most used estimators, the Horvitz-Thompson estimators (see chapter 5) of the totals are the same, or almost the same, as the true population totals for a set of control variables

b) *Random route sampling*: is a useful technique when access to register information of sufficient quality is not available. It is applied in CTS, but it is not advisable for ETS. Once randomly selected the first sampling unit, instructions are given to the interviewer to follow a random route and interview individuals, e.g.: take first road right, interview at second house on left, continue down the road, interview every tenth house on your right, etc.

c) *Systematic sampling*: the units are selected from a previously ordered list by randomly choosing the first element and then selecting the subsequent ones respecting a fixed interval proportional to the sample size. This technique maintains the probabilistic nature. Furthermore, being the starting list ordered according to some criteria, e.g. the Nomenclature of Territorial Units for Statistics (which is the acronym - from French - of: *Nomenclature d'Unités Territoriales Statistiques* – NUTS, see European Union, 2003) or its breakdown at local level in Local Administrative Units (LAUs), systematic sample also operates an implicit stratification of the units.

**(iii) Panel sample**

4.21 Usually, an ETS is used to keep track of changes in time of variables of interest. One of the reasons for this is that the results from an ETS form an early indicator for changes in the business cycle. Therefore, it is advantageous to choose a survey strategy in which changes in time are measured accurately.

To this aim, it is worth remembering an important feature the selected sample design can have: the panel feature. A panel can be any sample, whatever built, with the following characteristic: the same set of units is surveyed each month (or quarter). If all the units are maintained the sample is more properly called *fixed panel*, if part of units are periodically substituted, it deals of a rotating panel. Using a fixed panel rather than selecting a fresh sample each round of the survey, increases the precision of estimates of changes in time. The reason for this is that using a panel eliminates fluctuations in the results purely due to differences in the survey samples. The panel sample is therefore suited for use and widely applied in business surveys, less for consumers.

In CTS, in fact, the panel feature prevent from asking consumers for some sensitive questions as those required by the EC harmonised project (like income, or household economic situation). The surveyed consumer, in these cases, feel to be controlled and likely refuses to repeatedly respond these kind of questions.

In using a fixed panel in BTS, instead, there are great practical advantages because the initial contact with the enterprise – to determine the structure of the enterprise and agree on the reporting units – is time-consuming and therefore costly. However, once the same group of enterprises is surveyed in repeated rounds, it is no longer strictly random. This is because the target universe will change over time as new entrants appear and as existing enterprises cease trading or change their kind of activity.

4.22 To cope with the problem of changes in the target universe, some surveys are based on a *rotating panel* with a fixed percentage – say 25% – being replaced at regular intervals. A more common approach is to review the sample once a year with new enterprises brought in to replace those that have ceased operation or changed their activity. While there are administrative advantages for the organisers in maintaining the same enterprises in the sample for several rounds, there is a danger of "respondent fatigue". After too many questionnaires, respondents may refuse to reply or fail to give proper consideration to their answers.

The rotating panel design is also applied in the Surveys of Consumers of the Michigan University: within the 500 units composing the sample about 60% are new respondents, while 40% are interviewed for a second time, after a time interval of six months.

This design, according to Curtin (2013), has several distinct advantages over non-panel samples also for consumer surveys. From one point of view it allows to better detect the underlying causes of the changes highlighted by the panel structure. It also allows longitudinal researches, made possible by repeated measurements. It is possible to achieve larger samples by pooling of up to six of the independent monthly samples to carry out researches which need larger sample size, like those on rare population or events. As the interviewed units are recurrent, the non-sampling influences remain relatively constant across waves. It is worth remembering, however, that also if the panel has been initially randomly selected, it could not be considered a random sample in the subsequent waves.

**(iv) BTS: specific features of sample design**

4.23 An oft-used sample design for BTS is the stratified random sampling. One commonly stratifies according to properties of the units in the survey frame. Each stratum is then defined as the group of all units for which these properties are similar. The expectation is that units that have similar properties also have similar values of the target variable. As discussed above, stratifying in this manner gives more precise estimators. Three such properties are of particular interest to BTS:

- kind of Activity
- number of Employees (or another measure of size, e.g. sales)
- geographical location.

When stratifying one also needs to decide the depth of stratification. In other words: how similar must units in the same strata be, and how many different strata are allowed? Should we, e.g., distinguish just two categories for number of employees ("large" businesses and "small" ones), or should we use a more fine-grained stratification?

Which properties to use depends on a number of factors: obviously the required auxiliary information must be available for all units in the survey frame. Also, one's choice of unit will sometimes limit availability: for instance, geographical location cannot be used if the enterprise is used as the unit, because an enterprise may consist of establishments in different locations.

The output specification should be taken into consideration when dealing with these questions. If, e.g., output for different kinds of activity is required, it makes sense to stratify according to kind of activity. Moreover, it often does not make much sense to use a stratification scheme that is much more fine-grained than the output scheme.

One should also consider the resulting number of strata, especially when stratifying according to more than one property at the same time. If one, e.g., stratifies according to both Kind of Activity and Geographical location, and one distinguishes 30 kinds of activity and 40 economical regions, one ends up with 30 * 40 = 1200 strata. Such a high number of strata may well be prohibitive, given the maximum allowed sample size and the necessity for sampling at least a minimum number of units per stratum.

**(v) CTS: specific features of sample design**

4.24 The consumer sampling design usually stems from combination of different kinds of allocations techniques and selection of the units, therefore the consumer samples are usually featured by a complex sample design. Hereafter the main features are presented, while for a more detailed discussion see e.g. Cochran (1977), Särndal *et al.* (2003).

In some countries, for cost effectiveness and also for the advantages related on a well consolidated experience, the consumer survey rely on already existing social surveys so they make use - even with some adjustment - of the same samples.

It is important to establish how the sample has initially to be built, together with the definition of the size of the sample. These two aspects are strictly bound each other as both contribute to the desired precision of estimates (see chapter 5).

When selecting households as sampling units, some probabilistic methods have to be applied in selecting, within it, the individual who is going to be interviewed. One of the most applied criterion is the First (/last) Birthday criterion. The interviewer asks to contact the member of the household who firstly (/lastly) will have (/had) birthday. Another widely used technique is represented by the First Answer criterion. In this case the adult is interviewed who firstly answers the phone. This method is less expensive and leads to similar results to the previous one.

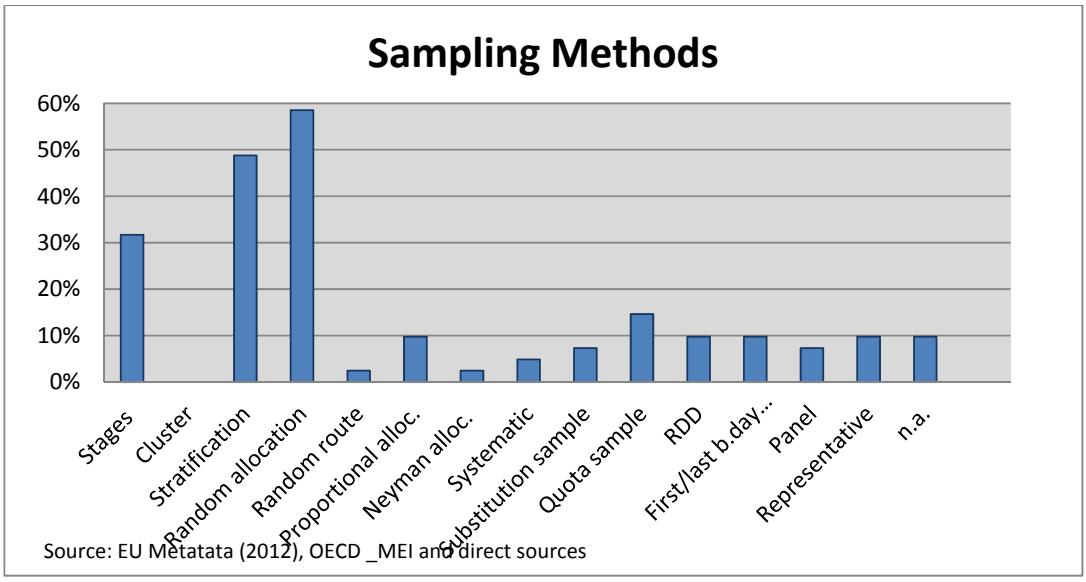
4.25 Most countries conducting tendency surveys have designed their sample to be representative in that sense that every adult member of the population has an equal chance of being selected: that is samples selected according to an Equal Probability of Selection Method (EPSEM).

One of the most widely used sampling techniques is the EPSEM clustered (multi-staged) stratified random sample. This kind of sample design is adopted in survey sampling for practical reasons. Mainly, if the sampling frame units cover two or more survey units (e.g. consumer within households), clustering is the only practical way of selecting a sample of the required units, or if we have to divide the sample into manageable workloads for interviewers, while using the face-to-face data collecting mode. The multi-stage feature, while allows saving of travelling time and consequent reduction in costs, however implies larger sampling error than un-clustered samples, as only a part of the population is monitored.

Nationally representative samples are standard in all developed countries; in emerging economies the samples may be limited to urban centres, such as in China, Malaysia, and Mexico, for example. Such limitations of the national samples are not critical omissions if the areas excluded from the sample do not fully participate in the market economies of their countries (Curtin, 2007).

**Figure F4.2**

**Sampling design features of Consumer Survey in different OECD, EU and EU accessing countries**




In Figure F4.2 the main features of the consumer design samples available for the OECD and EU and EU accessing countries are reported. Samples are built including several features and the large majority of the samples are complex, built as multistage stratified random samples. A lot of alternatives are however selected according to specificities and capabilities of the different countries. On average, all the samples are designed comprising at least two of the characteristic reported in the Figure.

4.26 An example of complex sample design for Consumer surveys is the ISTAT CTS sample. It is a stratified, two stages (telephone subscribers/consumers), random sample of exactly 2000 units, newly selected every month. The reference population is represented by the full aged (18+ years) population, stemming from Census survey and yearly updated with demographic statistics outcomes. The stratification is performed by grouping population in 42 strata according to six NUTS1 (Nomenclature of territorial units for statistics - level 1: partitions) and seven classes of demographic width of LAU2 (Local Administrative Units – level 2: municipalities).

The frame is made up of the telephone subscribers list, ordered by NUTS2 (regions), LAU2 and zip codes; the primary sampling unit is the telephone subscriber selected, within each stratum, with systematic random selection technique. The second stage sampling unit is the consumer. The consumer is selected within the adults belonging to the household which corresponds to the chosen telephone number. The selection criteria applied is the First Answer method, which is accomplished by taking in to account a quota system by gender and age. The inclusion probability is constant within each stratum with respect to the adult population.

It is worth noticing that the strata sample size exactly reflects the population proportion, but not that of the households' one. This occurrence affects the calculation of the inclusion probability (see chapter 5).

## References

American Association for Public Opinion Research (2011). *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys, 7th edition*. AAPOR. Available at: http://www.aapor.org/AM/Template.cfm?Section=Standard_Definitions2&Template=/CM/ContentDisplay.cfm&ContentID=3156.

American Association for Public Opinion Research (2013). *Report Of The AAPOR Task Force On Non-Probability Sampling*. Available at: http://www.aapor.org/AM/Template.cfm?Section=Reports1&Template=/CM/ContentDisplay.cfm&ContentID=5963.

American Association for Public Opinion Research, (2013). *Best Practices for Surveys and Public Opinion Research*. AAPOR, available at: http://www.aapor.org/Best_Practices1.htm .

Cochran, W. G. (1977). *Sampling techniques*. John Wiley & Sons.

Curtin, R. T. (2003). *Current Research and Development Agenda for the U. S. Consumer Surveys*. OECD workshop, Brussels, Belgium, November 21.

Curtin, R. T. (2007). *Consumer Sentiment Surveys: Worldwide Review and Assessment*. JBCMA, 2007/1.

Curtin, R. T. (2013). *Surveys of Consumers – Sample Design*. University of Michigan, last accessed May, 16 2013 at: http://www.sca.isr.umich.edu/documents.php?c=i.

European Commission (2006). *The Joint Harmonised EU Programme of Business and Consumer Surveys*. European Economy, Special Report n.5, ISSN 1684-033X http://ec.europa.eu/economy_finance/publications/publication7568_en.pdf .

European Commission, Directorate-General for Economic and Financial Affairs (2014), *"The Joint Harmonised EU Programme of Business and Consumer Surveys"*, user guide March, available at http://ec.europa.eu/economy_finance/db_indicators/surveys/method_guides/index_en.htm.

European Commission (2013). *COS/BTS Survey Metadata*. available at: http://ec.europa.eu/economy_finance/db_indicators/surveys/metadata/index_en.htm.

European Union (2003). *REGULATION (EC) No 1059/2003 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 26 May 2003 on the establishment of a common classification of territorial units for statistics (NUTS.* Official Journal of the European Union.

Eurostat (2003). *Business register Recommendations manual*.
Available electronically: http://epp.eurostat.ec.europa.eu/cache/ITY_OFFPUB/KS-BG-03-001/EN/KS-BG-03-001-EN.PDF.

Martelli, B.M. (2005). *Relationship between response rate and data collection methods – Analysis of key issues and main findings*. OECD Task force on "Improvement of Response Rate and Minimisation of Respondent Load", Joint European Commission – OECD Workshop on International Development of Business and Consumer Tendency Surveys, Brussels 14-15 November 2005 available at: http://www.oecd.org/dataoecd/55/40/35558806.pdf.

Särndal C. E.,Swensson B.,Wretman J. (1991-2003). *Model Assisted Survey Sampling*. Springer, New York.

University of Michigan (2013). *Survey of consumers*. Accessed on May, 2013 http://www.sca.isr.umich.edu/documents.php?c=iBanning, R., Camstra, A., Knottnerus, P., 2012, *Sampling theory: Sampling design and estimation methods*, Statistics Netherlands, The Hague/Heerlen. Available electronically: http://www.cbs.nl/NR/rdonlyres/092CE48D-7D83-4B56-82234EF9B9EFD7AB/0/2012Samplingtheoryart.pdf.

Knottnerus, P. (2011) *Panels: Business Panels.* Statistics Netherlands, The Hague/Heerlen.
Available electronically: http://www.cbs.nl/NR/rdonlyres/58E4FA6F-6802-4125-BE72-D3566ECD37B6/0/2011x3719.pdf.

OECD (2003). *Business Tendency Surveys: A handbook.* OECD Publications, Paris, ISBN 92-64-19894-6. (http://www.oecd.org/std/leading-indicators/31837055.pdf.

# CHAPTER 5

## Estimation procedures and accuracy

### A. Introduction

5.1 In this chapter, commonly used methods are presented to estimate values of the variables of interest from the data gathered in the sample survey. These methods, also called estimation procedures, must correct for any bias that arises, e.g., when the sample misrepresents the population as a whole. For this purpose, individual units in the sample are weighted, using four different types of weights:

- Size weights, used to deal with size differences between units (mainly used in BTS);
- Sample weights, which arise from the probabilistic nature of the sample;
- Post-stratification weights, which are related to the structure of the population (mainly used in CTS);
- Non-response weights, used to correct for non-response.

As the treatment of non-response is discussed in Chapter 7, the present chapter focuses on the three remaining kinds of weights.

5.2 This chapter also discusses the accuracy of different estimation procedures, and its relation with the population variance and the required sample size. As a general rule, accuracy improves when the population variance is smaller or the sample size gets bigger. General formulas to estimate the accuracy are presented and illustrated with a real-life example.

### B. Target variables and size weights

5.3 The term “target variable” is used for any variable for which a population estimate is desired. While an ETS generally has a number of different target variables, all of these can be treated similarly, so this section will not be concerned with the specifics of the variable chosen.

5.4 For CTS, the population target value is the simple arithmetic mean of the target variable at the unit level:

$$Y = \frac{1}{N}\sum_{i=1}^{N} y_i \qquad (1)$$

In this formula the following notation is used:

- $i$ denotes a unit in the population;
- $N$ denotes the total number of units in the population;
- $y_i$ denotes the value of the target variable for unit $i$;
- $Y$ denotes the population target variable;

5.5 As discussed in earlier chapters, ETS mostly deal with qualitative questions with three answer categories per question (positive, neutral, negative). For these questions there exists a separate unit target variable $y_i$ for each option, which means that every such question gives rise to three target variables. It can only have the values +1 (if the respondent chose that option) or 0 (otherwise). Other qualitative questions (which may have a different number of answer categories) can be dealt with in a similar manner. ETS can also contain quantitative questions; for each such question there is exactly one target variable $y_i$, with the value declared by the respondent.

5.6 For BTS, the situation is more complicated, because of size differences between units. Here, the population target variable depends as follows on the values of the target variable at the unit level:

$$Y = \frac{1}{W}\sum_{i=1}^{N} w_i y_i \qquad (2)$$

In this formula the same notation is used as before, with the following additions:

- $w_i$ denotes the weight of unit $i$. This weight is often called the size weight;
- $W = \sum_{i=1}^{N} w_i$ denotes the population weight: the total of the size weights of all units in the population.

In other words, for BTS the population target variable is the weighted mean of the target variable for all units in the population.

5.7 Note that formula 0 can also be used for CTS, by choosing $w_i = 1$ for all units. To avoid unnecessary duplication, size weights will therefore be included in all formulas in the rest of this

chapter. If one is specifically interested in CTS, one can simply remove $w_i$ from any formula and replace $W$ with $N$.

**(i) Why use size weights in BTS?**

5.8 Size weights are used in all variables typical of BTS, because the importance of the answers is assumed to depend on the size of the reporting units. The answers from a large firm carry more weight than answers from a small one.

5.9 Weighting is necessary, because the variables typically collected in BTS don't inherently represent the size of a business. "Ordinary", quantitative variables, on the other hand, do: broadly speaking, variables like turnover, numbers of employees and volume of production will be higher for larger businesses. But for the variables typically used in BTS this is not the case: they are either qualitative variables, or quantitative ones (such as the percentage of capacity utilisation) for which the values do not depend intrinsically on the size of the unit. Weighting can be seen as the first step of the quantification process applied to categorical data to obtain estimates.

**(ii) Choice of size weights in BTS**

5.10 Strictly speaking, the variables to be used as size weights should depend on the target variable concerned. For example, questions about production ought to be weighted by the relative value of production by a unit in the branch as a whole, questions about employment with the number of persons employed, etc. But it would be costly, or even impossible, to obtain such a set of weights for each reporting unit and survey variable. Furthermore, practical experience has shown that the target variables typically used in BTS are not very sensitive to the choice of weighting variables. Therefore, in practice it is sufficient to use a single variable reflecting the general economic importance of the enterprise in weighting all the survey answers.

5.11 If a single variable is used as size weight, value added is probably the best choice, because the BTS results will then most closely reflect movements in GDP. However, value added is typically not known for individual businesses. Therefore one will typically choose two different variables for the size weight: one variable, called the *inner weight* is used on the individual and branch levels (e.g. manufacture of footwear and clothing), and a different variable, called the *outer weight* is used when aggregating branches to higher levels (e.g. manufacture of consumer goods or total manufacturing).

5.12 Since value added is the best choice of weighting variable, it should be used as outer weight. A popular choice for the inner weight is the number of employees, as this is a good measure for company size, and within branches the number of employees and value added are often closely correlated. Other variables that are known on the unit level, such as turnover, can be used as well, as long as they are a good proxy for the outer weight.

5.13 To show how this works in practice, assume that the population $P$ is divided into mutually exclusive sub-populations $P_1, P_2, \ldots, P_M$, with $M$ the number of sub-populations. Then the population target variable can be determined by the following two-step process:

$$Y_j = \frac{1}{E_j} \sum_{i \in P_j} e_i y_i \qquad (3)$$

$$Y = \frac{1}{V} \sum_{j=1}^{M} V_j Y_j \qquad (4)$$

In these formulas we use the same notation as before, with these additions:

- $j$ denotes a sub-population;
- $e_i$ denotes the inner weight (e.g., numbers employed) of unit $i$;
- $E_j = \sum_{i \in P_j} e_i$ denotes the total inner weight of all units in sub-population $j$;
- $Y_j$ denotes the target variable for sub-population $j$;
- $V_j$ denotes the outer weight (e.g., value added) of sub-population $j$;
- $V = \sum_{j=1}^{M} V_j$ denotes the total of the outer weights for the entire population

It is easily shown that this two-step process is equivalent to using the following weights in equation 0:

$$w_i = e_i \frac{V_j}{E_j} \qquad (5),$$

with $j$ denoting the sub-population to which unit $i$ belongs. This can be seen as an approximation of the variable used as outer weight for unit $i$, using its inner weight as a proxy. For instance, applying numbers employed as inner weights and value added as outer weights gives the same results as approximating the value added of individual businesses by their numbers employed.

5.14 In order to simplify the presentation somewhat, no distinct inner weights and outer weights will be used in the rest of this chapter. Instead, the notation $w_i$ for the size weights will be used throughout. If a scheme with inner and outer weights is desired, equation 0 can be used to derive the size weights $w_i$ from the inner and outer weights.

## C. Estimators and sample weights

### (i) Horvitz-Thompson and ratio estimators

5.15 To estimate the population target variable based on the sample response, the standard and widely-used *Horvitz-Thompson (HT) estimator* (see, e.g., Cochran ,1977) or Särndal *et al.* (1993) for a thorough treatment; Banning *et al.* (2012) is a more accessible introduction) can be used:

$$\hat{Y}^{HT} = \frac{1}{W}\sum_{i\in p}\frac{w_i}{\pi_i}y_i \qquad (6)$$

Here the same notation is used as before, with the following additions:

- $\pi_i$ denotes the inclusion probability (i.e., the chance for a unit to be drawn in the sample) of unit $i$;
- $p$ denotes the set of units in the sample, which is a subset of the population. We denote the number of units in $p$ by $n$, henceforth called the sample size;
- $\hat{Y}^{HT}$ denotes the Horvitz-Thompson estimate of the population target variable $Y$.

Note that $1/\pi_i$ acts as a weight in this formula; it is called the *sample weight*. The values of $\pi_i$, and hence the sample weights, depend on the sample design. In subparagraph C(ii) formulas for $\pi_i$ are presented for the sample designs discussed in the previous chapter.

5.16 Another widely-used estimator is the *ratio estimator*. It can be obtained from the formula for the HT estimator by replacing $W$ with its HT estimator $\hat{W}^{HT}$:

$$\hat{Y}^{R} = \frac{1}{\hat{W}^{HT}}\sum_{i\in p}\frac{w_i}{\pi_i}y_i \qquad (7), \text{ with}$$

$$\hat{W}^{HT} = \sum_{i\in p}\frac{w_i}{\pi_i} \qquad (8).$$

Here $\hat{Y}^{R}$ denotes the ratio estimate of the population target variable $Y$.

5.17 Both the HT estimator and the ratio estimator have their pros and cons. The HT estimator is unbiased[14], which means that the estimate is, on average, equal to the target variable, whereas the ratio estimator isn't. However, it can be proved that the bias of the ratio estimator is small if the sample size is large enough: under the "rule of 30" it can usually be neglected.

5.18 The ratio estimator, on the other hand, is often more precise than the HT estimator, in particular if $y_i$ and $w_i$ are correlated and the spread (measured by its variance) of $w_i$ is small. Because of this, the HT estimator is usually the best choice for a PPS sampling design, while the ratio estimator may be a good candidate for a stratified sampling design.

5.19 Another difference, which is mainly relevant to BTS, between the HT estimator and the ratio estimator is that the former assumes *a priori* knowledge of the size weights of all units in the frame, whereas the latter only needs the weights of the units in the sample. This means that the ratio estimator can be used if the size weights of the units are *a priori* unknown, simply by adding a question to the questionnaire about the variable used as size weight (e.g., number of employees).

5.20 Note that 0, 0 and 0 are only valid for the entire population. In case one is also interested in the target variables for specific sub-populations, one can easily modify these formulas to obtain the estimators for the target value for some subset $Q$ of the entire population (where $Q$ is not necessarily equal to the $P_j$ introduced earlier):

$$\hat{Y}_Q^{HT} = \frac{1}{W_Q} \sum_{i \in q} \frac{w_i}{\pi_i} y_i \; ; \tag{9}$$

$$\hat{Y}_Q^{R} = \frac{1}{\hat{W}_Q^{HT}} \sum_{i \in q} \frac{w_i}{\pi_i} y_i = \sum_{i \in q} \frac{w_i}{\pi_i} y_i \Big/ \sum_{i \in q} \frac{w_i}{\pi_i} \tag{10}$$

Notation is as before, with these additions:

- $q$ denotes the set of units in $Q$ that are also in the sample;
- $W_Q = \sum_{i \in Q} w_i$ is the total weight of all units in sub-population $Q$ ;

[14] Strictly speaking, the HT estimator is an unbiased estimator for the population target variable only if the survey frame and the population are identical. As we have seen in the previous chapter, in practice this is not the case. In this chapter it will be assumed that bias due to differences between the survey frame and the population is negligible.

- $\hat{W}_Q^{HT}$ is the HT estimator of $W_Q$;
- $\hat{Y}_Q^{HT}$ and $\hat{Y}_Q^R$ respectively denote the HT estimator and ratio estimator of the target variable for sub-population $Q$.

**(ii) Inclusion probabilities**

5.21 In order to use the estimators introduced in the previous subparagraph must know the inclusion probability for each unit in the population. In this subparagraph we therefore discuss the inclusion probabilities of the sample designs introduced in the previous chapter.

5.22 For *SRS*, the inclusion probability is the same for all units: $\pi_i = n/N$.

5.23 For *Stratified Sampling*, we can use this result to obtain the inclusion probabilities for units in stratum $h$: $\pi_i = n_h/N_h$, in which $n_h$ denotes the number of units in the sample and $N_h$ the number of units in the population in stratum $h$. The reason for this is that within each stratum, SRS is used.

5.24 With the *proportional allocation method*, $n_h$ is determined as follows:

$$n_h = \frac{N_h}{N} n \qquad (11).$$

Hence the inclusion probability simplifies to $\pi_i = n/N$, the same as for SRS. Note, however, that this does not mean that SRS and Proportional Allocation are equivalent: compared to SRS, some extreme samples (e.g., ones in which all sample units lie in the same stratum) are excluded.

Sample designs in which all inclusion probabilities are the same are called *self-weighting*. For self-weighting designs the sample weights don't need to be calculated separately, and the formulas for the estimators are greatly simplified.

5.25 For *Neyman allocation*, the following formulas are used:

$$n_h = \frac{N_h S_h}{\sum_{h=1}^{H} N_h S_h} n \qquad (12);$$

$$S_h^2 = \frac{1}{N_h - 1}\sum_i \left( w_i y_i - \frac{1}{N_h}\sum_i w_i y_i \right)^2 \quad (13).$$

Here the summations run over all units within the stratum. The following notation is introduced:

- $h$ denotes a stratum;
- $H$ denotes the number of strata;
- $S_h^2$ denotes the adjusted stratum variance of the "weighted" target variable $w_i y_i$ for stratum $h$. $S_h^2$ is a measure of the homogeneity of a stratum: if it is low, the stratum is homogeneous; if it is high, the stratum is inhomogeneous, meaning that there is considerable variability in the value of the target variable for different units.

5.26 In *PPS sampling* the self-selecting units must be treated separately. For these units one simply finds $\pi_i = 1$, because they are always included. For the non-self-selecting units, the following holds:

$$\pi_i = \frac{w_i}{W'} n' \qquad (14)$$

Here $n'$ denotes the reduced sample size one gets when the self-selecting units are excluded: $n' = n - n_s$, with $n_s$ the number of self-selecting units. Likewise, $W'$ denotes the total weight of all units in the population, with the exception of the self-selecting ones. Denoting the set of self-selecting units by $s$, we get $W' = W - \sum_{i \in s} w_i$.

5.27 If one substitutes 0 in 0 or 0 it becomes apparent that the size weights disappear from the formulas. This means that in PPS sampling, the *un-weighted* sample mean is used to estimate a *weighted* population mean.

5.28 A *two-stage design* is, in effect, a combination of two sample designs, one for each stage. In the first stage, clusters of units within the population are selected, and in the second stage, individual units are selected from each cluster. This means that the inclusion probabilities of a two-stage design can be obtained by combining the inclusion probabilities of both stages:

$$\pi_i = \pi_c^{(1)} \pi_{i,c}^{(2)} \,. \qquad (15)$$

Here we use the following notation:

- $c$ denotes the cluster in which unit $i$ resides;
- $\pi_c^{(1)}$ denotes the probability with which cluster $d$ is selected in the first stage of the sampling process;
- $\pi_{i,c}^{(2)}$ denotes the probability with which unit $i$ is selected from cluster $d$ in the second stage of the sampling process.

5.29 Any combination of sample designs discussed in this subparagraph (including the two-stage sample design itself: this gives a design with three or more stages) can be used to construct a two-stage sample design. Therefore, the number of possible two-stage designs is very large, and we will limit ourselves to discussing a single, real-life example.

5.30 Consider the case where, for CTS, consumers are selected by first selecting households, and then selecting one eligible individual from each household. Households are selected using a stratified sampling design (stratified, e.g., according to region), and a person is selected from each household using SRS. By using the formulas for the inclusion probabilities of these designs, introduced earlier in this subparagraph, we find:

$$\pi_c^{(1)} = \frac{n_h}{F_h}, \quad \pi_{i,c}^{(2)} = \frac{1}{f_c}, \qquad (16)$$

in which

- $i$ denotes an eligible person;
- $c$ denotes the household to which person $i$ belongs;
- $h$ denotes the stratum in which household $c$ resides;
- $n_h$ denotes the number of persons selected in stratum $h$ (which is equal to the number of households selected in the stratum);
- $F_h$ denotes the number of households in the population in stratum $h$;
- $f_c$ denotes the number of eligible persons in household $c$.

From this the overall inclusion probability to select person $i$ can be determined:

$$\pi_i = \frac{n_h}{F_h f_c}. \qquad (17)$$

Compare 0 with the inclusion probabilities of a stratified sampling design with proportional allocation, with $n_h$ consumers per stratum. While the latter design is self-selecting, the current one is not, due to the varying number of eligible individuals within each household.

5.31 The following table summarises the inclusion probabilities for all sample designs discussed in this subparagraph:

**Table T5.1**
**Inclusion probabilities for different sample designs**

| Sample design | Inclusion probability |
|---|---|
| SRS | $\pi_i = n/N$ |
| Stratified Random Sampling: Proportional allocation | $\pi_i = n/N$ |
| Stratified Random Sampling: Neyman allocation | $\pi_i = \frac{S_{h(i)}}{\sum_{h=1}^{H} N_h S_h} n$ |
| PPS (self-selecting units) | $\pi_i = 1$ |
| PPS (other units) | $\pi_i = \frac{w_i}{W'} n'$ |
| Two-stage design | $\pi_i = \pi_d^{(1)} \pi_{i,d}^{(2)}$ |

In this table the following notation is introduced:

- $h(i)$ denotes the stratum to which unit $i$ belongs;

**D. Post-stratification**

5.32 Post-stratification is a technique, mainly used in CTS, to apply a correction to the estimated value when the sample is not an accurate reflection of the population. One cause for such inaccuracies is when values of stratification variables such as age, gender or education level of sample units are not known until after data have been collected. In such cases one may find that some categories are over- or under-represented, simply because of the random nature of the samples used.

5.33 The data collection mode can also contribute to this. For example, when conducting a consumer (not household) survey by telephone, extreme age classes may be under-represented: young people are more difficult to contact while the elderly may be less confident in participating.

Something similar may happen for gender or occupation: women are often easier to contact than men, and working people are more elusive, as the time they spend home is limited.

5.34 In order to avoid introducing a bias in the results due to this kind of misrepresentation, a correction can be applied to the sample weights to ensure that the sizes of sub-populations are estimated accurately. To do this, the population $P$ is first divided into mutually exclusive sub-populations $P_1, P_2, \ldots, P_M$ (often called post-strata), with $M$ the number of sub-populations. Then the post-stratification estimator is as follows:

$$\hat{Y}^P = \frac{1}{W}\sum_{j=1}^{M}\frac{W_j}{\hat{W}_j^{HT}}\sum_{i\in p_j}\frac{w_i}{\pi_i}y_i, \qquad (18)$$ with

$$\hat{W}_j^{HT} = \sum_{i\in p_j}\frac{w_i}{\pi_i}. \qquad (19)$$

Here we have used the same notation as before, with a couple of additions:

- $\hat{Y}^P$ denotes the post-stratification estimate of the population target variable $Y$;
- $p_j$ denotes the set of units in the sample that belong to post-stratum $P_j$;
- $W_j = \sum_{i\in P_j} w_i$ denotes the total weight of all units in $P_j$, which is assumed to be known (perhaps from a different survey);
- $\hat{W}_j^{HT}$ denotes the Horvitz-Thompson estimate of $W_j$.

The ratios $W_j / \hat{W}_j^{HT}$ act as weights in equation 0; these weights are called post-stratification weights. They can be interpreted as modifiers to the sample weights $1/\pi_i$. The method itself is more properly called complete post-stratification, as a complete knowledge of $W_j$ is assumed for all post-strata.

5.35 To obtain more insight into this abstract formula, we present the post-stratification estimator for the simplest case: SRS without replacement while ignoring size weights. Then equation 0 simplifies considerably:

$$\hat{Y}^P = \sum_{j=1}^{M}\frac{N_j}{N}\left(\frac{1}{n_j}\sum_{i\in p_j} y_i\right), \qquad (20)$$

with the same notation as before, with these additions:

- $n_j$ denotes the number of elements in $p_j$;

- $N_j$ denotes the number of elements in $P_j$.

In contrast, the Horvitz-Thompson estimate for this case, rearranged for easy comparison, is:

$$\hat{Y}^{HT} = \sum_{j=1}^{M} \frac{n_j}{n} \left( \frac{1}{n_j} \sum_{i \in p_j} y_i \right), \qquad (21)$$

In other words, both estimators first take the arithmetic mean over the sample units in each post-stratum, and then average these over the post-strata. However, the post-stratification estimator uses the relative *population* sizes of the post-strata as weights, whereas the Horvitz-Thompson estimator uses the relative *sample* sizes instead. As a consequence, if the sample is skewed with respect to the population, the Horvitz-Thompson estimator may yield estimates that are less accurate than the post-stratification estimator.

5.36 An different technique from the one outlined above, called Ratio estimation, requires that an auxiliary variable is known for each post-stratum. More precisely this technique assures that, for a given auxiliary variable the estimated total equals the given value (Vanderhoeft, 2001).

5.37 When only the marginal distributions of the auxiliary variables are known but the cross classification cell counts are lacking or unreliable, or the size of the cells is extremely small, other approaches must be applied. These methods are called incomplete post-stratification methods, of which these are the most widely used:

- The Regression estimator, which introduces multiple post strata indicator variables;
- The Raking ratio estimator, or iterative proportional fitting method, in which weights are computed to satisfy marginal constrains in a cross tabulation;
- The Calibration method, which can be regarded as a more general method, with ratio, regression, and raking ratio estimators as special cases.

5.38 The principle underlying the Calibration method, proposed by Deville and Särndal (1992), see also Sautory (1993, 2003), is to adjust samples trough modifying unit weights using auxiliary information stemming from a set of available information referred to as calibration variables.

The adjustment is made by replacing the original sample weights with new weights so that:

- For a categorical (qualitative) variable the proportions values of the replay options estimated within the sample, after re-weighting, will become the same of the corresponding values known for the population;

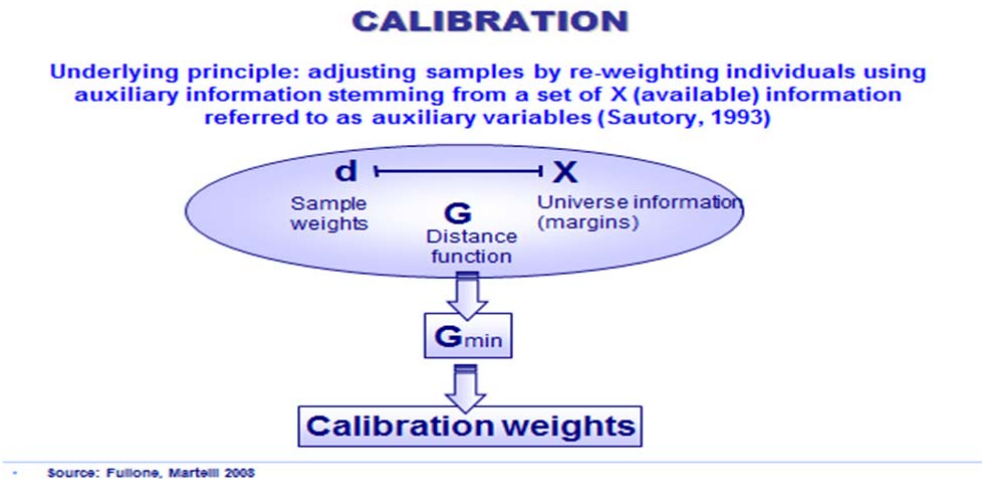
- For a numeric (quantitative) variable, the total of the estimated variable within the sample, after re-weighting, will be equal to the known total of the population.

This weighting method allows for reducing the sampling variance and in some cases to lower the bias due to the unit non-response.

In figure F5.1 a graphical representation of the method is presented.

**Figure F5.1**

**Graphical representation of the Calibration method**




5.39 The Calibration estimator can be formulated in a similar way as the Horvitz Thompson estimator and can be written as:

$$\hat{Y}^{CAL} = \frac{1}{W}\sum_{i \in p} v_i y_i \,. \qquad (22)$$

Here notation is the same as before, with one addition:

- $v_i$ denotes the calibration weight of unit $i$, which takes into account both the size weights (if present) and the sample weights, and the values of selected auxiliary population variables $X$ on which the sample is weighted.

The Calibration method finds, with an iterative process, the values which minimize the selected distance function G. Even if the calculations are rather complex, some programs are freely

available on internet for processing the calibration weights. See e.g. CALage sur MARges (CALibration on MARgines - CALMAR) of INSEE[15] and RGENESEES of ISTAT[16].

**Box B5.1**

**CALMAR in practice**

In the following, a brief description is given of the application of one of them, namely of CALMAR procedure, as recommended by the EC, to the Italian CTS. For an easier application, that is for shortening the iterative process of convergence, CALMAR requires as starting weights the effective sampling weights, which are calculated according to the previous section. As calibration variables, besides the regional ones, those that usually are more affected by the sampling design are considered. The selected calibration variables then more precisely are: Age (4 classes), NUTS2 (19 Italian Regions), LAU2 (7 classes of municipalities according to size), Occupation (4 categories); Education (3 categories). We excluded gender (as already considered in the quota interviews). The aim was to balance as much detail as possible with a reasonable convergence over time.

These variables are calculated for every year of the considered time span (total populations, regional population, age, occupation), or as often as they became available (for education, stemming from censuses, only a ten-year revision was performed). It was not possible to use income as the population data are not quickly available and reliable. As the selected population variables are not available at cell level, more properly an incomplete post-stratification was performed.

Out of the four possible models (distance functions) offered by CALMAR for calculating weights, the logit method was selected, as the linear ones admit negative values and the raking ratio allows quite high extreme values. This method is based on a logistic distance function and needs to have the domain (lower and upper bounds) of the function defined. This setting is rather subjective. A procedure was set-up that, for every month, first applied the raking ratio method. The resulting bounds identified by raking ratio were used as starting point for setting the domain of the logit

[15] The SAS macro CALMAR was developed by the French National Institute of Statistics and Economic Studies (INSEE) (http://www.insee.fr/en) and is available at: http://www.insee.fr/fr/methodes/default.asp?page=outils/calmar/accueil_calmar.htm

[16] The software: R Evolved Generalised Software for Sampling Estimates and Errors in Surveys was developed by The Italian National Institute of Statistics, Italy (www.istat.it/en/) and is available. at: http://www.istat.it/it/strumenti/metodi-e-software/software/regenesees

method, gradually decreasing the upper (and increasing the lower) bound *t* times as to reach values for which convergence was no more reachable. Then we considered the *t*-1 attempt the desired one.

In the Italian experience the lower (*L*) and upper (*U*) margins resulted in the whole time span considered, on average, *L*=0.41 and U =1.55. These bounds resulted rather constant over time: in the 1995-1997 their averages were equal to *L*=0.33 and *U*=1.51, may be, in this span, the less efficient sample design being compensated by the collapsing of regional margins into partitions. In the 2004-2006 period they resulted *L*=0.44 and *U*=1.56. The final calibrated weights thus resulted ranging from about 2,500 to 37,000, being a composite result of the initial weights and of the post stratification process. That is every interview, depending on composition of household and features of the respondent, represents a number of the total population ranging from 2.5 thousand up to, for the most "rare" cases, nearly 40 thousand individuals. The resulting weights, once divided by their average are associated with each interview, in this way summing up *n*=2000 (the sample size).

**E. Accuracy and sample size**

5.40 Since the sample does not include all members of the population, estimates based on sample data generally differ from parameters of the entire population. The difference between the sample estimate and the population true value is defined sampling error. The sampling error is a measure of the precision of estimates.

5.41 Exact measurement of sampling error is generally not feasible since the true population values are unknown and is due to the fact that only a subset of the population is monitored; however, sampling error can often be estimated by the probabilistic nature of the sample.

**(i) General theory**

5.42 Our purpose is to define a sample size $n$ so that the estimate $\hat{Y}$ of the unknown population variable $Y$ lies within an accepted interval around $Y$ with a low probability $\alpha$ to commit an error. That is:

$$\Pr(|\hat{Y} - Y| \leq d) = 1 - \alpha \,. \qquad (23)$$

Here $d$ denotes the desired precision of the estimate, that is the maximum allowable difference (or error) from the true value $Y$ we accept in calculating the estimate.

5.43 If the sample is large enough $\hat{Y}$ has an approximately normal distribution with variance $\text{var}(\hat{Y})$. If we assume that any bias in the estimate is negligible, the normal distribution is centred on $Y$. Hence the ratio

$$Z = \frac{\hat{Y} - Y}{\sqrt{\text{var}(\hat{Y})}} \tag{24}$$

has a standard normal distribution, and

$$\Pr(|\hat{Y} - Y| \leq d) = \Pr\left(|Z| \leq \frac{d}{\sqrt{\text{var}(\hat{Y})}}\right) = 1 - \alpha . \tag{25}$$

From the tables of the standard normal distribution we know that the above relation is true if

$$\frac{d}{\sqrt{\text{var}(\hat{Y})}} = z_{\alpha/2}, \tag{26}$$

with $z_{\alpha/2}$ the abscissa that cuts off an area of size $\alpha/2$ from the right tail of the standard normal distribution. Note that, in effect, an area of the same size is also cut off from the left tail of the distribution, as we are dealing with the absolute value of $Z$, which leaves an area of size $1-\alpha$.

5.44 For the HT estimator, $\text{var}(\hat{Y})$ can be determined as follows:

$$\text{var}(\hat{Y}^{HT}) = \frac{1}{W^2} \sum_{i \in P} \sum_{j \in P} \frac{\pi_{ij} - \pi_i \pi_j}{\pi_i \pi_j} w_i w_j y_i y_j . \tag{27}$$

In words: the variance of the Horvitz-Thompson estimator is a linear combination of products of the form $y_i y_j$, for all pairs of units in the population. Here the same notation is used as before, with one addition:

- $\pi_{ij}$ is the probability that units $i$ and $j$ are both part of the sample. This parameter is also called the "second order inclusion probability".

Note that in practice, this formula can be hard to use, because it presumes a knowledge of $y_i$ for all units in the population. If values of $y_i$ are only known for the units in the sample, the following estimator can be used:

$$\hat{\text{var}}(\hat{Y}^{HT}) = \frac{1}{W^2} \sum_{i \in p} \sum_{j \in p} \frac{\pi_{ij} - \pi_i \pi_j}{\pi_{ij} \pi_i \pi_j} w_i w_j y_i y_j . \tag{28}$$

This estimator is sometimes called the Horvitz-Thompson estimator of the variance. It is a linear combination of products $y_i y_j$ for all pairs of units in the sample.

5.45 For SRS, simple expressions for $\pi_{ij}$ can be derived:

$$\pi_{ii} = \frac{n}{N}, \text{ and } \pi_{ij} = \frac{n(n-1)}{N(N-1)} \text{ for } i \neq j. \qquad (29)$$

For most other sample designs the exact expressions for $\pi_{ij}$ are difficult to derive, making it hard to use variance formulas 0 and 0. Reasonable approximations for the variances are available in certain situations, but it would go beyond the scope of this document to give these expressions for $\pi_{ij}$ for all the sampling designs presented in this Handbook. The interested reader is referred to the standard literature (see, e.g., Särndal et al., 1993).

5.46 For the other estimators presented in this Handbook, approximate formulas for the precision can be derived, but these are even more complicated than for the HT estimator. As before, interested readers are referred to the standard literature. Suffice to say that the ratio estimator is often more precise than the HT estimator, in particular if there is a correlation between the size weight and the target variable.

**(ii) Example: Simple Random Sampling**

5.47 While the formulas for the sample variance given in the previous subparagraph are valid for all sample designs, they do not offer much insight in the accuracy of particular sampling strategies. However, in the case of a Simple Random Sample, formula 0 simplifies considerably:

$$\text{var}(\hat{Y}^{HT}) = \frac{N}{W^2}\frac{N-n}{n}S^2. \qquad (30)$$

Here $S^2$ is the adjusted population variance of the "weighted" target variable $w_i y_i$:

$$S^2 = \frac{1}{N-1}\sum_{i\in P}\left(w_i y_i - \frac{1}{N}\sum_{j\in P} w_i y_i\right)^2. \qquad (31)$$

The parameter $S^2$ is a measure of the variability of the weighted target variable within the population. If $S^2$ is low, the population is homogeneous, i.e., the value of the weighted target variable is almost the same for all units. In contrast, if $S^2$ is high, the weighted target variable varies considerably from unit to unit.

5.48 Equation 0 shows that the accuracy of the estimator increases with increasing $S^2$, and decreases with increasing sample size. In other words: the estimator is more precise if the sample is larger or if the population is more homogeneous. This matches our intuition: If we were to survey all units in the frame (in which case $n = N$), the estimator would give the exact value of the target variable, whereas if the value of the weighted target variable were the same for each unit (in which case $S^2 = 0$), we would only need to survey a single unit to know the value of this variable for all units, and hence obtain the exact value of the population target variable.

5.49 If we ignore size weights and use the fact that $y_i$ can only attain the values 0 and +1, , that is in the case of ETS categorical data, we can derive a very simple expression for $S^2$:

$$S^2 = \frac{N}{N-1} Y(1-Y) . \qquad (32)$$

Combining this with 0 and 0 we finally get the following relation between $d$, the precision of the estimate and $n$, the sample size:

$$\frac{d^2}{z_{\alpha/2}^2} = \frac{N-n}{N-1} \frac{Y(1-Y)}{n} . \qquad (33)$$

In this, the ratio $(N-n)/(N-1)$ is called the finite population correction. Since the sample is often much smaller than the population, this correction can usually be ignored. We then get

$$d = z_{\alpha/2} \sqrt{\frac{Y(1-Y)}{n}} , \qquad (34)$$

or, solving for $n$,

$$n_0 = \frac{z_{\alpha/2}^2}{d^2} Y(1-Y) . \qquad (35)$$

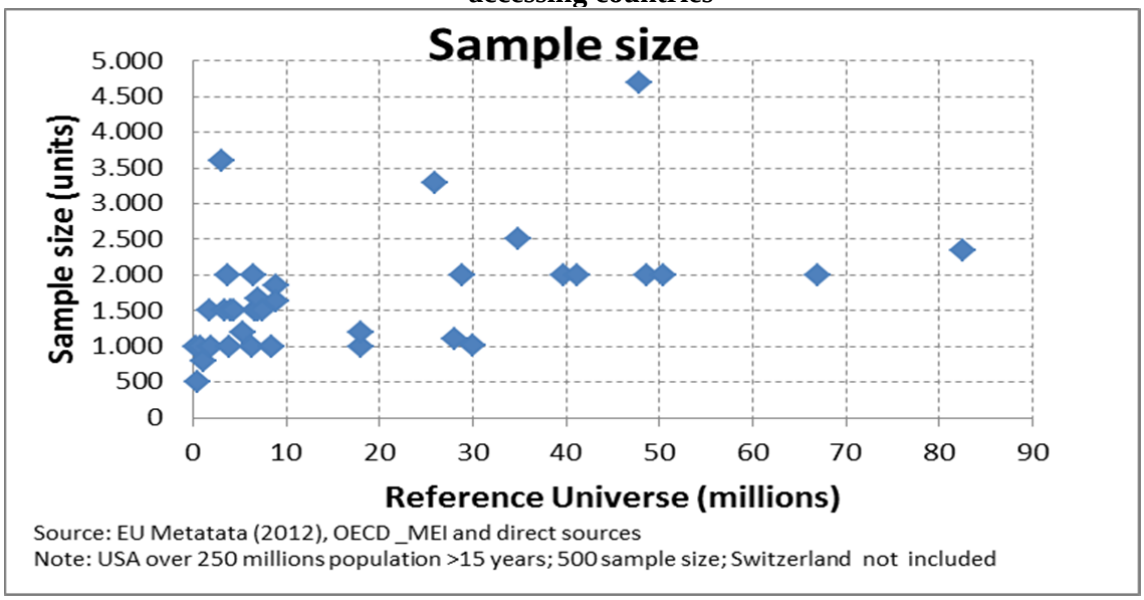
In practice we first calculate $n_0$. If $n_0 / N$ is negligible, $n_0$ is a satisfactory approximation of $n$. If not, $n$ can be obtained from

$$n = \frac{n_0}{1 + (n_0 - 1)/N} . \qquad (36)$$

5.50 An important consequence of 0 is that in large populations the sampling size is relatively independent from the population size. That is, the sample estimates are not bound to the size of the population they represent. As a matter of fact, the majority of the countries carrying out a CTS,

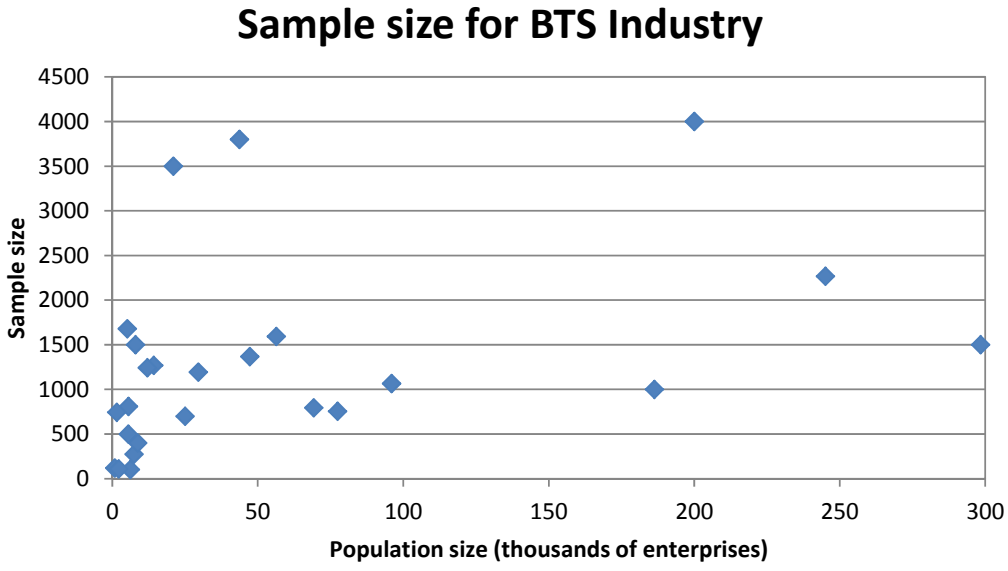
select a sample size ranging from 1000 to 2000 units, from populations running in the millions, as illustrated in Figure F5.2. Similar sample sizes are used in BTS (industry only), as can be seen from figure F5.3.

**Figure F5.2**
**Sample size of CTS as a function of population size for different OECD, EU and EU accessing countries**




Source: EU Metatata (2012), OECD _MEI and direct sources
Note: USA over 250 millions population >15 years; 500 sample size; Switzerland not included

**Figure F5.3**
**Sample sizes of BTS (industry only) as a function of population size for EU countries**




Source: EU metadata (2012). Romania not included.

5.51 For example, let us choose $d = 0.03$ as the margin of error we accept for our estimates, and $\alpha = 0.05$, which means we are willing to accept a small risk of 5 per cent that the actual error is larger than the margin of error. The value of $n_0$ also depends on the (generally unknown) value of $Y$, but in the case we don't have any further information, we can assume the situation of maximum uncertainty, given by $Y = 1 - Y = 0.5$. Then $z_{\alpha/2} \approx 1.96$ and we obtain:

$$n_0 \approx \frac{1.96^2}{0.03^2} 0.5 * 0.5 \approx 1067. \qquad (37)$$

5.52 As in CTS the population is generally large (say $N > 100{,}000$) the finite population correction does not give any significant contribution. In fact, using 0, the result remains nearly the same:

$$n \approx \frac{1067}{1 + (1067 - 1)/100{,}000} \approx 1056. \qquad (38)$$

5.53 In BTS, however, population for small countries can be fairly small. If that is the case, the finite population correction gives a significant contribution. If, say, the number of businesses in industry is 8000, equation 0 gives the following result:

$$n \approx \frac{1067}{1 + (1067 - 1)/8{,}000} \approx 942. \qquad (39)$$

5.54 Table T5.2 shows the outcomes for simple random samples according to the formula 0 for different values of $Y$ and different desired precision. It is worth noticing that, $d$ being constant, the sample size doesn't vary dramatically in the interval between 0.3 and 0.7. On the other hand the sample size depends on $d$ in a disproportionate way while trying to increase precision (that is to lower $d$). E.g. if we wish to pass from $d = 3\%$ to $d = 2\%$ we have to more than double the sample size, and from $d = 2\%$ to $d = 1\%$ we have to enlarge the sample by nearly four times.

**Table T5.2**
**Sample size as a function of target variable and precision**

| Margins of error | $\alpha = 5\%$ | $1-\alpha = 95\%$ | $z_{\alpha/2} \approx 1.96$ |
|---|---|---|---|

| | **Target variable** $Y$ | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Precision** $d$ | *0,1* | *0,2* | *0,3* | *0,4* | ***0,5*** | *0,6* | *0,7* | *0,8* | *0,9* |
| *0,01* | 3457 | 6146 | 8067 | 9220 | 9604 | 9220 | 8067 | 6146 | 3457 |
| *0,02* | 864 | 1537 | 2017 | 2305 | 2401 | 2305 | 2017 | 1537 | 864,3 |
| ***0,03*** | 384 | 682,9 | 896 | 1024 | **1067** | 1024 | 896 | 682,9 | 384,1 |
| *0,04* | 216 | 384,1 | 504 | 576 | 600 | 576,2 | 504 | 384,1 | 216,1 |
| *0,05* | 138 | 245,9 | 323 | 369 | 384 | 368,8 | 323 | 245,9 | 138,3 |
| *0,06* | 96 | 171 | 224 | 256 | 267 | 256 | 224 | 171 | 96 |
| *0,07* | 71 | 125 | 165 | 188 | 196 | 188 | 165 | 125 | 71 |
| *0,08* | 54 | 96 | 126 | 144 | 150 | 144 | 126 | 96 | 54 |
| *0,09* | 43 | 76 | 100 | 114 | 119 | 114 | 100 | 76 | 43 |

5.55 A further important component has to be considered too: the cost of the survey. Maybe the obtained sample size is too high for the available budget so that it is necessary to accept a larger margin of error (i.e., by increasing $d$). In that case we can use equation 0 to estimate the level of precision that can be attained. Table T5.3 shows the best attainable precision according to this equation for different values of $Y$ and different sample sizes, again with a confidence interval of 95%. It is evident that the errors do not vary significantly for values of $Y$ between 30 and 70 per cent.

**Table T5.3**

**Precision as a function of sample size and target variable**

**(percentage points)**

| Margins of error | $\alpha = 5\%$ | $1-\alpha = 95\%$ | $z_{\alpha/2} \approx 1.96$ |
|---|---|---|---|

| | **Target variable** $Y$ | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Sample size** $n$ | *0,1* | *0,2* | *0,3* | *0,4* | *0,5* | *0,6* | *0,7* | *0,8* | *0,9* |
| *100* | 5,9 | 7,8 | 9,0 | 9,6 | 9,8 | 9,6 | 9,0 | 7,8 | 5,9 |
| *200* | 4,2 | 5,5 | 6,4 | 6,8 | 6,9 | 6,8 | 6,4 | 5,5 | 4,2 |
| *300* | 3,4 | 4,5 | 5,2 | 5,5 | 5,7 | 5,5 | 5,2 | 4,5 | 3,4 |
| *400* | 2,9 | 3,9 | 4,5 | 4,8 | 4,9 | 4,8 | 4,5 | 3,9 | 2,9 |
| *500* | 2,6 | 3,5 | 4,0 | 4,3 | 4,4 | 4,3 | 4,0 | 3,5 | 2,6 |
| *750* | 2,1 | 2,9 | 3,3 | 3,5 | 3,6 | 3,5 | 3,3 | 2,9 | 2,1 |
| *1000* | 1,9 | 2,5 | 2,8 | 3,0 | 3,1 | 3,0 | 2,8 | 2,5 | 1,9 |
| *1250* | 1,7 | 2,2 | 2,5 | 2,7 | 2,8 | 2,7 | 2,5 | 2,2 | 1,7 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| *1500* | 1,5 | 2,0 | 2,3 | 2,5 | 2,5 | 2,5 | 2,3 | 2,0 | 1,5 |
| *2000* | 1,3 | 1,8 | 2,0 | 2,1 | 2,2 | 2,1 | 2,0 | 1,8 | 1,3 |
| *2500* | 1,2 | 1,6 | 1,8 | 1,9 | 2,0 | 1,9 | 1,8 | 1,6 | 1,2 |
| *3000* | 1,1 | 1,4 | 1,6 | 1,8 | 1,8 | 1,8 | 1,6 | 1,4 | 1,1 |

**(iii) Design effect**

5.56 The design effect is a measure of the efficiency of a sample design. It is defined as the variance of the estimate of the design, relative to the variance of the estimate obtained by SRS:

$$Deff = \frac{\text{var}(\hat{Y})}{\text{var}(\hat{Y}_{SRS}^{HT})}. \qquad (40)$$

If the design effect of a given design is less than 1, the estimate is more precise than the corresponding SRS design, meaning that the design is more efficient, and hence the sample can be smaller while maintaining the same accuracy.

5.57 Kish (1965) verified that, comparing the variances of estimates obtained from a simple random, a stratified random sample with proportional and Neyman allocation, we obtain:

$$\text{var}(\hat{Y}_{NEY}^{HT}) \leq \text{var}(\hat{Y}_{PROP}^{HT}) \leq \text{var}(\hat{Y}_{SRS}^{HT}). \qquad (41)$$

In other words, Neyman allocation is more efficient than proportional allocation, which in turn is more efficient than SRS.

It is worth remembering, however, that if the variances of the target variable are very low, even the most efficient complex sample designs offer only a moderate increase in efficiency compared to the simple random sample.

5.58 As an example, in the Italian CTS, with a sample of size of 2,000, the average sampling error for the questions composing the confidence indicator, at a 95% confidence interval, is on average not greater than 1.4%. Thus $Deff < 1$, indicating that the selected sample design has a higher precision than the simple random sample.

## References

Afkhami, R., Crockett, A. (2005). *Weighting the Social Surveys*. ESDS Government, available at www.esds.ae.uk.

Banning, R., Camstra, A., Knottnerus, P. (2012). *Sampling theory: Sampling design and estimation methods*. Statistics Netherlands, The Hague/Heerlen. Available electronically: http://www.cbs.nl/NR/rdonlyres/092CE48D-7D83-4B56-8223-4EF9B9EFD7AB/0/2012Samplingtheoryart.pdf.

Cochran, W. G. (1977). *Sampling techniques*. John Wiley & Sons.

Deville, J. C., Särndal C. E. (1992). *Calibration Estimators in Survey Sampling*. Journal of the American Statistical Association, June, Vol. 87, n:418, pp376-381.

European Commission (2006). *The Joint Harmonised EU Programme of Business and Consumer Surveys*. European Economy, Special Report n.5, ISSN 1684-033X http://ec.europa.eu/economy_finance/publications/publication7568_en.pdf.

European Commission, Directorate-General for Economic and Financial Affairs (2014), *"The Joint Harmonised EU Programme of Business and Consumer Surveys"*, user guide March, available at http://ec.europa.eu/economy_finance/db_indicators/surveys/method_guides/index_en.htm.

Fullone, F., Martelli, B.M. (2008). *Re-thinking the ISAE Consumer Survey Processing*. ISAE working paper no. 92, Rome. Available at: http://ideas.repec.org/p/isa/wpaper/92.html.

Kish L. (1995). *Survey Sampling*. Wiley Classics Library, New York.

Malgarini, M. (2005). *Efficient sample design and weighting methods-Analysis of key issues and recommendation.* OECD Taskforce on "Harmonisation on Survey Operation and Technical Design", Joint European Commission –OECD Workshop on International Development of Business and Consumer Tendency Surveys, Brussels 14-15 November.

OECD (2003). *Business Tendency Surveys: A handbook.* OECD Publications, Paris, ISBN 92-64-19894-6. (http://www.oecd.org/std/leading-indicators/31837055.pdf.

Särndal, C.-E., Swensson, B., Wretman, J.H., 1993. *Model Assisted Survey Sampling.* Springer-Verlag, New York, ISBN 0-387-97528-4.

Sautory, O. (1993*). La macro CALMAR_ Redressement d'un échantillon par calage sur marges*. Document n. F9310, INSEE.

Sautory, O. (1993). *CALMAR (Calage sur marges)*. www.insee.fr.

Vanderhoeft, C. (2001). *Generalised Calibration at Statistics Belgium*. Statistic Belgium Working Paper N. 3, NIS.

# CHAPTER 6

# Data collection

## A. Data collection: designing a strategy

### (i) Primary data collection

6.1 Primary data collection involves gathering of statistical data through various means, one of which is through a survey. Before undertaking any survey, there is a need to have an overall strategy for data collection for a more organized and systematic data collection system. Thus, for ETS, it is important that there is data collection strategy.

6.2 The data collection strategy will strongly depend on the available financial and manpower resources of the implementing agency. As much as possible, it is hoped that the objectives of the survey are met given the available resources. Designing data collection strategy may also depend on some key elements, such as the variables to be collected or how the questions were formulated, the channels and modes to be used for communicating with the respondents, and the channels or modes to be used to collect the information.

6.3 The design of the strategy for data collection has direct influence on the quantity and quality of the input for the data processing and on the necessary steps to reduce non-sampling errors (chapter 7). A well-designed data collection strategy has the following benefits: collecting more information of higher quality, motivating the respondents to cooperate, yielding high response rate, and minimizing non-sampling errors.

6.4 Before one starts collecting data through a survey on a regular frequency, a well- designed data collection strategy is needed. A data collection strategy consists of two elements: the technique used to send and receive the information (‘mode’) and communication with the respondent. A good strategy helps in structuring the process of data collection: it minimizes measurement and non-response errors and maximizes willingness to cooperate and response rate. It also makes data collection cost efficient

**(ii) Data collection strategy: survey techniques and communication with respondents**

6.5 The different survey techniques use different technical tools for collecting data. Due to technological progress, the number of 'modes' for collecting data has increased. A distinction can be made between interviewer assisted modes and self-completed modes in which the respondent completes the survey without assistance. A combination of modes is sometimes used to optimise the data collection (mixed-mode approach). In subparagraph B(ii) the available modes will be described.

6.6 The second aspect is communication with respondents. Good communication helps to convince the respondent to participate, to eliminate misinterpretation of the request of information and eventually to counter problems in the receiving of the information. Communication with respondents occurs during pre-survey stage, during field work stage or in between surveys. In subparagraphs B(i) and B(iii) these aspects will be described.

6.7 When combining the two aspects in a data collection strategy, it is possible to use a different mode for the communication than for the actual data collection. This is called a mixed- or multimode survey system (Biemer, Lyberg, 2003).

**(iii) Determinants in designing the process of data collection**

6.8 When designing a data collection strategy, the process will be influenced by various aspects. First, the scope of the survey must be considered, what is the survey about and what population is being researched. Second, quality aspects play an important role. The different sources of errors must be limited. Finally some organisational aspects have impact on the data collection strategy.

*Scope of the survey*

6.9 ETS can be either conducted among businesses or consumers. For data collection, this has consequences. Usually, businesses are more complex to observe, due to heterogeneity of businesses. It can be more difficult to find the right contact person than with households. On the other hand, high level employees of businesses are better capable of completing questionnaires without help, while consumers usually need more clarification on the questionnaire.

6.10 BTS are usually conducted among a fixed sample of enterprises (a fixed panel), while CTS usually are conducted among a new sample each time. These different kinds of samples need a different approach in data collection. When dealing with a fixed panel, investing in communication

with the respondent is more important. Once a good relationship has been established, a steady response is more likely and the respondent needs less assistance answering the questions. When dealing with onetime respondents, more effort in motivating to comply is required and probably some assistance during the answering is needed.

6.11 The design of the questionnaire also has impact on the data collection strategy. The kinds of questions asked and how much assistance and clarification is needed determines if an interviewer assisted or self-completed mode is preferred. The length of the questionnaire and the expected duration of answering also influence the survey technique and the level of communication needed to persuade respondents to comply. ETS are generally intuitively answerable and the questionnaire is fairly short.

6.12 The questionnaire determines the choice for the data collection mode, but once the mode is selected, the questionnaire has to be aligned to suit perfectly for the selected mode.

*Quality aspects*

6.13 The operation of data collection is affected by measurements errors. Measurement error is the difference between a measured value of quantity and its true value (Dodge, 2003). In surveys, these errors can be mistakes in the collection of data, including both the incorrect recording of a response and the correct recording of a respondent's inaccurate response. Measurement errors are related largely to the questionnaire. Measurement errors in ETS are largely avoided by use of internationally harmonised questionnaires.

6.14 Other measurement errors are related to the data collection strategy. Mode effects occur when modes provide different outcomes with the same questionnaire. This can be the case for example with sensitive topics. Interviewer assisted modes may lead to socially acceptable answers. Selection effects can occur when a certain mode excludes a part of the panel, for example because a part of the panel has no registered telephone line or access to internet. Memory effects occur when respondents cannot remember correctly the past events. When a former answer is given within the new questionnaire, this can influence the 'memory' of the respondent.

6.15 Another kind of errors derive from processing the statistical data. This kind of error is introduced when the collected data have to be edited or have to be entered manually into a processing program. If the questionnaire is in the form of computer programme, processing errors

can be dramatically reduced. The computer automatically poses the questions in the right sequence and verifies the answers immediately. Checks can be built in to correct answers that are outside of the acceptable range.

*Organizational aspects*

6.16 One unique feature of ETS is its timeliness. The results are often published immediately after the reporting period (mostly monthly). The total processing period is limited to a maximum of 20 working days. The data collection period is often limited to 10 to 15 days. This very short collection period limits the use of some survey techniques, while it benefits others.

6.17 An important factor is the cost of conducting ETS. This includes wages, material, IT equipment, time, processing time, etc. A complex data collection strategy can lead to a high response rate, but will cost more than a simple approach. There is a trade-off between available resources and the desired quality of the collected data.

6.18 There is political pressure to reduce the administrative burden on respondents (especially businesses). A data collection strategy must make it very easy to comply for the respondent, and not over-asking the respondent. Questionnaires must therefore be as user-friendly and short as possible. Sampling of course helps to minimize the administrative burden, but samples should not be unnecessary large. Furthermore a rotating panel can diminish the burden on businesses.

## B. Designing the process of data collection

### (i) Pre survey: preparations and communication

6.19 Pre-survey preparations and communication is another important aspect in data collection. Usually a letter informing about the upcoming survey and citing the objectives of the survey is sent to the target respondents, especially for BTS. Another way of communicating about the survey is through a press release or announcement to be provided to the media or to be posted on the website of the implementing agency. This form of communication would be appropriate both for business and consumer tendency surveys.

6.20 For BTS, there is a need to identify the contact persons in order to facilitate the data collection process. The list of firms or enterprises from which the sample would be drawn usually contains information such as the name of the firm or enterprise, address, major activity, size of the

company, and contact information. Initial coordination with the sample firms or enterprises should be done in order to get the name of the contact persons.

6.21 For CTS, the sample households is usually drawn from available list of households. The list of households usually contains information such as the name of the household head and address. Initial coordination with the sample households should be done in order to facilitate the data collection. The head of the household or any responsible member of the household is usually the respondent. If the survey is carried out by telephone technique, initial coordination is not necessary.

**(ii) Field work: collecting the data**

6.22 The first distinction between the different survey techniques available is whether the respondent is assisted by an interviewer or completes the questionnaire on his own. The form in which the questionnaire is presented and recorded is the second distinction.

6.23 A personal face to face interview, nowadays assisted by a computer, is labelled with the abbreviation CAPI (Computer Assisted Personal Interviewing). An interviewer reads out the questions and notes down the answers. The interviewer and respondent are able to communicate verbally and nonverbally. It is possible to offer extra information, like brochures, to clarify the questionnaire. For complex questionnaires, some sort of routing can be used. The interviewer can intervene in case of misunderstanding or incorrect answers from the respondent.

6.24 The second form of a personal interview is over the telephone, labelled CATI. This form of data collection is mostly organised from an internal or external call centre. The interviewers can centrally be instructed, guided and supervised. The costs are relatively high, although not compared to CAPI, and the interview can be done fast. This form of interview is less suited for long and complex questionnaires. However, for CTS, this is quite an effective mode.

6.25 Telephone and face to face interviews avoid some forms of measurement errors and the issue of non-response. With these modes, the required information is immediately and fully available. The interviewer assures that all questions are answered as intended. The experience with this type of observation is that the personal approach usually provides a higher response rate than self-completed methods.

6.26 The complete and correct information and a higher response rate because of the interviewer come with a price tag. The cost of the personal observation is much higher than other modes,

because of the staff costs. This kind of observation also takes a long period of fieldwork. Also the burden on respondents is rather high: an interviewer comes to their home or company and the visit is relatively time consuming.

6.27 For ETS, a very short data collection period is essential. The economic conditions can change every day, thus influencing the survey results in case of big events during the data collection phase. Furthermore, the questionnaire itself is rather simple, so in most cases, assistance from an interviewer is not necessary.

6.28 The traditional way of conducting surveys is with paper questionnaires. The respondent receives the questionnaire by mail and completes it without help of an interviewer. The questionnaire is then sent back by mail. The disadvantage of this type of observation is that the back and forth sending of postal questionnaires yields a delay of at least two days. An alternative is using e-mail for sending and returning the questionnaire, in that case there is less delay. This mode is not only faster than sending questionnaires by post, but also cheaper. Both postal and e-mail questionnaires pose a risk of incomplete answers: partial / item non-response. Also, questionnaire by postal mail usually yields low response rates.

6.29 A newer form of data collection is via the internet. On a secure website a questionnaire can be placed and visited by the respondent. The respondent can answer in a quick and efficient way and it is also possible to incorporate some automatic assistance (sometimes referred to as Computer Assisted Self-Administrated Questionnaire or CASAQ). The advantages of this mode are largely comparable to the other forms of self-completed data collection because it is both a fast and relatively inexpensive way to collect data for the ETS. Another advantage of this mode is that completing all the questions can be enforced so that item non-response is not an issue. Quality checks on the answers are also an option. Hard checks prevent the respondent from continuing until a probable answer has been given. Soft checks just display a warning to the respondent to indicate an improbable answer. Analysts and processors prefer questionnaires with many checks. However, one should be cautious to use too many automatic checks because they can irritate respondents and might cause non-response if they quit answering halfway the questionnaire. Furthermore improbable answers may actually reflect the real world.

6.30 A combination of these modes can be used in a mixed-mode approach. Using different modes simultaneously gives an opportunity to compensate for the weaknesses of each individual

mode. For the respondent it can be a reason to cooperate when he can choose his preferred mode to his convenience.

6.31 Especially for ETS, a mixed mode approach can contribute to a higher response rate in the short period of data collection. Furthermore, with a mixed mode approach, respondents unreachable by a certain mode, can be reached with another mode, thus eliminating some selective under coverage.

6.32 The different modes of data collection each have their advantages and disadvantages. On the basis of the considerations, a choice will have to be made for the preferred mode. Table T6.1 shows the main advantages and disadvantages mentioned for each mode.

**Table T6.1**

**Advantages and disadvantages of data collection modes**

| Mode | | Advantages | Disadvantages |
|---|---|---|---|
| Interviewer-administered | Face to face (CAPI) | • interviewers can motivate respondents<br>• completed questionnaires<br>• possibility to explain and clarify questions<br>• high quality data<br>• permits lengthy interviews | • expensive<br>• socially desirable answers<br>• complexity, long fieldwork period |
| | Telephone (CATI) | • central instruction and supervision possible<br>• timeliness<br>• completed questionnaires<br>• possibility to clarify questions | • inappropriate for lengthy interview<br>• hasty answers<br>• socially desirable answers<br>• relatively expensive |
| Self-completed | Paper questionnaire (or questionnaire sent and received by e-mail) | • sensitive subjects can be researched<br>• available to respondents with no internet access<br>• sending and returning is cheap | • item non-response<br>• limitations to complexity of the questionnaire<br>• trouble reaching the right contact person in businesses<br>• delay due postal<br>• data entry, processing and correcting can be costly |
| | Web forms | • timeliness<br>• low cost<br>• completed questionnaires> 'hard' and 'soft' checks for improbable answers | • access to internet is needed<br>• low motivation of respondents<br>• low response if survey is not mandatory |

| Mode | | Advantages | Disadvantages |
|---|---|---|---|
| | | • automated processing | |
| | Mixed Mode | • high response rate, reduced (selective) non-response<br>• timeliness | • costs<br>• mode effect |

**(iii) Response enhancing measures**

6.33 The aim of data collection is to collect enough data so that the survey can give correct estimates. An important aspect is about the measures to generate the desired level of response. Response enhancing measures can be divided in two groups: proactive measures, during the preparation of data collection (during the development of the data collection strategy) and reactive measures when during the data collection the response is unsatisfactory.

6.34 The proactive approach uses persuasion principles in the development of letters, questionnaires, folders, interview protocols and websites, in order to convince potential respondents of the importance and necessity of responding. There is a distinction between pre-survey communication and field work communication. Pre-survey communication focuses on the initial respondent contact, introducing the survey and seeking possible cooperation. Field work communication is concerned with the questionnaire and the associated chosen style of communication.

6.35 If the response is unsatisfactory regardless of the mode of data collection, the statistician will be obliged to take measures (e.g. sending reminders) in order to achieve the desired result, albeit belatedly (reactive). This implies that follow-up steps will be needed even after applying the proactive response-enhancing measures with care. These follow-up steps are a part of the field work.

*Proactive*

6.36 To enhance response, one should try to adopt the point of view of the respondent. Participation in surveys costs businesses' and consumers' time, and with no gains in return, completing questionnaires has a low priority for most respondents. Thus, the significance of the survey should be clear, as well as of the importance of the particular respondent's participation.

6.37 The quality of the provided material influences response rate considerably. A good questionnaire should only ask for relevant information, should be attractive, and should be well-designed. Harmonised Business and Consumer Surveys generally meet the quality standards. Assistance of a person from the survey group should be available if necessary. A friendly tone usually helps in maximizing response rates. Furthermore the language used to communicate can be adjusted to the respondent. Cooperation with trade organisations and industry associations can help to establish good relationships with responding businesses.

6.38 Timing of the questionnaires and reminders is important, because the respondents must have the requested information to be able to comply properly. As a proactive measure, an announcement can be sent to prepare the respondents for the survey.

*Reactive (reminder policy)*

6.39 For some respondents, participation in a survey depends on whether completing the questionnaire is mandatory or voluntary. The response rates in voluntary surveys are generally lower than in studies involving a legal obligation to cooperate. A good reminder policy is important especially for the non-compulsory business and consumer tendency surveys to achieve high response rates. For reminders to non-respondents, the same modes of communication for data collection (see section C) may be used. However, it could be beneficial to use also a different mode with respect to the original one. The reminder policy should be based on costs, speed and effectiveness. If reminders are sent too quickly, this may annoy the respondent.

**Box B6.1**

**Data collection strategies used by Statistics Netherlands**

Statistics Netherlands (SN) uses mixed-mode survey systems in the ETS. The actual data collection is done in a single mode, but in communicating with the respondent, more modes are used. SN applies different strategies for the BTS and the CTS.

*BTS*

For the BTS, SN uses web forms as the data collection mode. The choice of this mode is a result of weighing the pros and cons of the different modes. The survey is conducted with a fixed panel of companies, so the person providing the data is familiar with responding through the web form. The advantage of web forms is that they are quick and cheap. An additional advantage is that answering

all questions can be enforced, so there is no issue with item non-response. Furthermore, the data can be processed automatically.

Communication starts well before the data collection period, which is conducted in the first two weeks of the reporting month. Before the first questionnaire, new companies receive an introduction letter, in which the survey is announced. Companies are asked to provide a contact person in advance, so that the questionnaire can be directed to the right respondent.

When the data collection period is opened, the companies receive a request by letter or e-mail to fill in the web form on a secured area on the website of SN. Along with the request, they receive a login code and password to gain access to the restricted area, where the survey can be completed. The login code is changed for each survey.

SN deals with unit non-response to the BTS using the same mode as in the data collection. The non-respondents receive a reminder sent by post or e-mail including the login code for the secure login. Some non-respondents are more reluctant to respond than others, so if necessary, several reminders are being sent. The procedure is to send businesses a first reminder about a week after the initial request to provide data, if they have not yet responded. If a business fails to respond again, a second reminder is sent a week later.

In practice, these actions are sufficiently effective to achieve the target response rate of 80 percent. On an occasional basis, one or two times a year, a telephone campaign is carried out to contact long-term non-respondents. This reminder campaign is expensive, but desirable because of the results, with the benefits outweighing the costs. Furthermore, SN has a service centre available for respondents, so that they can contact SN for questions, exchange contact information and report their complaints, etc.

In the past, SN used a mixed mode for data collection. The respondent could choose between a paper questionnaire and the web form. Additionally, during a telephone reminder session, respondents were given the chance to complete the questionnaire. This process was quite costly and time consuming. With fairly little impact on the response rate, this strategy was replaced with the current approach of web form only.

*CTS*

For CTS, the chosen observation mode is the personal interview by phone (CATI) and the panel changes monthly. Because of the higher response risk and given the limited size of the panel, SN

has decided to interview the consumers in person, by telephone. This approach has, as already indicated, the big advantage that the response rate is much higher than with other modes. With telephone interviews, a reminder policy for non-response is not needed. In the first contact with the consumer, it immediately becomes clear if that person may or may not wish to participate in the study. The households receive a letter in advance with the request to participate in the survey.

**Box B.6.2**

**Data collection strategies used by the Philippines**

In the Philippines, ETS are conducted by the Central Bank of the Philippines (Bangko Sentral ng Pilipinas). These surveys are carried out every quarter.

*BTS*

For BTS, the following modes of data collection are used: mail, telephone or e-mail. After determining the final list of the samples from the latest list of Top 7,000 Corporations of the Securities and Exchange Commission, letters and questionnaires are sent to the firms included in the samples. Target response rate for business surveys is 70 to 75 percent. Reminders and follow-ups are done by telephone or email. In the case of non-responding firms, replacement of samples is usually done to maintain the desired sample size by sector.

*CTS*

For CTS, data collection is done using only one mode which is the face-to-face interview method. Before the start of the field work, a field Operations Manual is prepared and a training for the survey supervisors and enumerators is provided. The head of the household or any responsible member of the household serves as the survey respondent.

Considering that personal (face-to-face) interview is adopted, a higher response rate of 95 percent is targeted for CTS. With this mode of data collection, follow-ups are very minimal or not necessary at all.

## C. Organizing the data collection

### (i) Training of enumerators

6.40 In organizing the data collection, training is an important component but the training will greatly depend on what modes to be used for data collection. After the survey questionnaire has been finalized, the data collectors need to be familiarized with the questions or items in the questionnaire. The training will also help in the identification of critical points to be highlighted in the survey and in the clarification of some concerns that may arise.

6.41 For developed countries which make use mostly of the web or internet as the mode for data collection, training is not really a concern. The training of enumerators is more appropriate for personal interview method, which is used by most of the developing countries. In the case of ETS, considering that the questionnaire is not that long, it can be gauged that the training required for the enumerators will not be so complicated.

**(ii) Technical facilities and documents**

6.42 Technical facilities such as the required computer specifications or programs for data collection should be made available. In order to avoid problems in collecting the data, especially for the web-based data collection system, there is a need to define the requirements or specifications for the technical facilities needed. System testing is also necessary before the actual conduct of data collection.

6.43 For personal interview method, having field operations manual or enumerator's manual would definitely help to organize and implement better the data collection system.

## References

Berkenbosch, B., (2011). *Response-enhancing measures for business surveys.* Statistical methods (201109), Data collection strategy, the Hague/Heerlen 2011, Statistics Netherlands.

Beukenhorst, D., Kerssemakers, F.,(2012). *General description and unimode designs.* Statistical methods (201211). Data collection strategy, the Hague/Heerlen 2012, Statistics Netherlands.

Biemer, P.P. and Lyberg, L.E. (2003). *Introduction to Survey Quality*. New York, John Wiley.

Dodge, Y., (2003). *The Oxford Dictionary of Statistical Terms*. OUP.

De Leeuw. E. D.. *To Mix or Not to Mix Data Collection Modes in Surveys*. Journal of Official Statistics, Vol. 21, No. 2, 2005, pp. 233–255.

Luiten, A. (2011). *Response-enhancing measures for social statistics.* Statistical methods(201121). Data collection strategy, the Hague/Heerlen 2011, Statistics Netherlands.

# CHAPTER 7

## Managing sources of non-sampling errors

### A. Non-sampling errors

7.1 The total error which affects estimates obtained through a sample survey can be due to *sampling* errors and *non-sampling* errors. The first component arises because almost all current statistical surveys measure the target variables on a subset of units belonging to the whole population. Thus, the outcome from the interviews is based on a selection of units and not from the entire universe. If unbiased estimators are used, the sampling error is equal to the sample variability of estimates, which could be measured through the sample variance of the estimator. If the survey is based on a random sample, errors arising from this source will be stochastic and the mean of the errors drawn from repeated samples will be zero.

7.2 Non-sampling errors arise from many sources including defects in the sampling frame because the business register is incomplete or out of date, improper selection of the units to be sampled, refusal from some selected units to provide information (total or partial refuse), and mistakes made at the stages of collecting and editing the answers or entering them into the data base (codification, registration, revision). Since there is not a complete and unique theory able to face all the possible non-sampling errors, the best practice consists in building up the "error profile" (description of the main sources of non-sampling errors) for each specific survey. Non-sampling errors may arise even when the whole reference population is observed. They may produce bias, meaning that the estimated values may be systematically higher or lower with respect to true values. Normally, measurement of non-sampling errors is quite difficult and strictly depends on the specific kind of survey dealt with.

7.3 As regards a generic target variable $y$ (which in this contest is a certain modality concerning a given opinion on tendencies) the total Mean Squared Error (*MSE*) error, given by the sum of sampling and non-sampling errors, is:

$$MSE = \sigma^2 + B^2 \qquad (1)$$

where $\sigma^2$ is the variance of the estimates for the universe based on a random sample and *B* is the bias of the estimate. If random sampling is used, an estimate of $\sigma^2$ can be computed from the sample. The bias is the deviation between the true value and the expected value of the estimates and contains the net effect of all the non-sampling errors mentioned above. It is not possible to measure the size of the bias, but the risk of errors can be reduced by testing the measurement procedure, maintaining a reliable business register and keeping non–response to a minimum. If random sampling is not used, there often will be no $\sigma^2$ component in the *MSE*, but the risk of bias will be increased and the total *MSE* may be larger than if random sampling had been used.

7.4 On the basis of a simple schematization, the various kinds of errors which may occur in the process of a statistical survey can be resumed as follows:

a) Sampling error
b) Non-sampling errors
    i) Coverage
    ii) Measurement and processing
    iii) Non-response
        - Total
        - Partial

**(i) Coverage errors**

7.5 The target universe is determined by specifying which groups of units are to be included in terms of:

- the characteristics of the unit (for business: kind of activity, juridical status, size and location; for household: kind of municipality, number of components);
- the period during which reporting units must be active in order to be included in the survey.

7.6 Ideally, each survey should cover all units which are "alive" at some time during the period(s) covered in that survey.

There are five main blocks of possible coverage errors:

a) *Not completeness*: population includes some units which do not belong to the list or to the business register
b) *Clusters of units*: the same name in the list is associated to more than one unit in the population

c) *Unknown or not existing names*: the list contains some names that do not correspond to any unit in the population

d) *Over-coverage*: the list contains names which are contained but the unit does not belong to the population

e) *Replicated names*: the population includes units to which correspond more than one name in the list.

The main consequence of these errors is that they influence the real inclusion probabilities respect to the original sampling design.

7.7 In the case of *not completeness*, the population is not completely included in the list (example: fixed telephone lists). We indicate as $y$ the target variable, as $P_L$ the relative frequency of units correctly included in the list $L$ (and as 1-$P_L$ the relative frequency of those not included); $\bar{y}_L$ and $\bar{y}_{\bar{L}}$ are, respectively, the mean of $y$ in the two parts of the population, so that the overall population mean $\bar{y}$ can be written as:

$$\bar{y} = \bar{y}_L P_L + \bar{y}_{\bar{L}}(1 - P_L) \qquad (2)$$

If we estimate $\bar{y}$ using the estimator $\hat{\bar{y}}_L$ which is an unbiased estimator of the subpopulation mean $\bar{y}_L$, the bias of the estimator $\hat{\bar{y}}_L$ is given by:

$$E(\hat{\bar{y}}_L) - \bar{y} = \bar{y}_L - \bar{y} = (1 - P_L)(\bar{y}_L - \bar{y}_{\bar{L}}) \qquad (3)$$

The bias depends on the share of units not included in the list and the difference between the means in the two subpopulations.

Possible approaches to reduce bias are as follows:

- we can create separate strata for units not included in the list (if we know that these units exist) and interview them using, for instance, an area sampling.
- we can integrate more than one list (combination among lists) in order to cover with the final list quite all the population units.

7.8 In the case of *clusters of units*, if clusters are rare and small, we can include in the sample all units belonging to the clusters, in order to give to each unit the same inclusion probability than the name drawn from the list (example: the name is an address, the units are all the households living at the same address). We can lose efficiency if clusters include too homogeneous units.

Otherwise, we can choose at random only one unit from the cluster and assign to it an inclusion probability inversely proportional to the cluster size. The consequence is that the sampling design is not self-weighting.

A third possibility consists in selecting a wider sample, choosing a sub-sample of units having the desired size.

7.9 *Unknown, or not existing names or over-coverage.* We can think to household who moved out, or firms that stopped or changed their activity. In this case, it is necessary: 1) to identify not eligible units; 2) to decide if accepting or not a sample size lower than that desired. If we want to carry out substitutions, we could estimate in advance the rate $\delta$ of not existing names and then select a sample of size $n/(1-\delta)$ in order to achieve to a sample with size $n$. Otherwise, if we want to use a substitution list, it must be created according to the same sampling scheme used for the selection of the first sample.

7.10 *Replicated names*. We can imagine an enterprise whose local units are included in the list with distinct names, so that the unit is present in the list more than once.

Generally, the inclusion probability is given by the sum of all the inclusion probabilities of the corresponding labels in the list. If one uses a self-weighting design, then we can use the ratio estimator:

$$\hat{\bar{y}}_Q = \frac{\sum_{i=1}^{n} y_i / A_i}{\sum_{i=1}^{n} 1 / A_i} \qquad (4)$$

where $n$ is the number of respondent units, $A_i$ is the number of labels in the list associated to the *i-th* unit. This result (which could lead to a less biased estimator) is due to the fact that the inclusion probability of a population unit is $\pi A_i$, where $\pi$ is the constant inclusion probability (under the self-weighting hypothesis) of labels. Since the probability to find more than once the same unit in the same sample is quite low, it is not convenient to limit actions to eliminate duplications from the sample.

*The case of telephone surveys*

7.11 When telephone consumer surveys are carried out, normally the list used for drawing the sample is the national list of household which have a fixed telephone number. Of course, such kind of list is affected by various "coverage" problems, among which under-coverage and over-coverage

are the most dangerous phenomena. Examples of over-coverage are telephone numbers which correspond to second houses and professional activities; main cases of under-coverage are given by families which do not have a fixed telephone, or which have a fixed telephone but do not want to be identified through that (for instance, they may have confidential telephone numbers). Moreover, some families have more than one number referred to the main dwelling, or there may be telephone numbers to which more than one household correspond.

7.12 Each case above leads to uncontrolled modification of the original inclusion probabilities defined through the sampling design, and may produce bias in final estimates. In particular, under-coverage will produce bias as much higher as more different will be the statistical profiles of units belonging or not to the telephone list. On the other hand, over-coverage will lead to an effective sample size lower than the desired one, which will imply sampling variance increase. A simple way to tackle the problem – as already mentioned – consists in incrementing the sample size for taking into account non eligibility of some sample units, or in using substitute lists.

7.13 As regards bias, it should be always evaluated *ex post*. If some structural variables are available for all units in the population (belonging or not to the list; for instance: kind of municipality, age, sex), the comparison between average profiles of units belonging to the effective sample and to the whole population may provide useful information on potential bias, provided that the available variables are correlated with the main target variables. Moreover, in theory a certain *ad hoc* correction of estimates may be applied, in order to recover bias due to under-coverage and/or other potential causes. Another possibility consists in finding and using a second list, which will partially overlap with the first one, but will also contain new population units not belonging to the first list. Using both lists for drawing a combined sample may reduce potential bias, but it is worthwhile to remark that it is necessary to carry out a preliminary evaluation of the bias affecting the second list: if the second list is affected by a bias larger than that of the first list and with the same algebraic sign, it may be better to use the first list only.

7.14 A generalized, well known tool for tackling bias in sampling surveys is given by the recourse to calibration estimators (Sarndal and Lundström, 2005, pp.179-190), which are founded on the idea to modify the original sampling weights in order to find new weights which have the property, when applied to one or more auxiliary variables, to reproduce given population totals, available from other sources or from estimates. Calibrated estimates should have a lower bias than not calibrated ones, even though the recourse to new sampling weights different from the original

weights will lead to a parallel increase of sampling variance. The balance between these two counterbalancing effects should be always carefully evaluated.

**(ii) Measurement and processing errors**[17]

7.15 These errors derive from the survey process as a whole. There may be a difference between the true value of a characteristic and that measured in the survey context. The difference may derive from the respondent himself (lack of precision, lack of information, deliberate wish to provide wrong data), or it may be due to some elaboration steps following the data capturing phase: for instance, that is the case of errors occurring during data registration, codification of open responses, manual identification and correction of values erroneously supposed to be wrong. Some of the problems may be reduced adopting modern tools for obtaining and saving responses (electronic questionnaires, via web or CAPI, CATI techniques[18]), which may already include some basic checks regarding measurement errors.

7.16 A feature of the most successful ETS is that the survey team maintains regular contact with the respondents by telephone or e-mail. Personal contacts of this kind increase response rates, and can also be used to improve the survey by identifying questions that respondents do not properly understand, cannot easily answer or questions that are not seen as relevant by the respondents. In addition, personal contacts can provide feedback on how the survey can be made more useful to the respondents and on how publications of the results can be improved. On the other hand, CTS should preferably avoid frequent contacts with respondents, who are often rotated at each survey occasion: the "surprise" effect characterizing new entries into the sample may be a key issue in order to achieve to high reliable responses.

7.17 Wrong responses are object of re-interview or, more frequently, imputation. The most recommended strategy for managing editing and imputation in a categorical data context is still represented – as well as when quantitative data are concerned – by the Fellegi-Holt method (De Waal et al., 2011, pp. 115-129).

*Measurement Errors*

7.18 Measurement errors occur when the observed value is different from the true one. It can be due to:

a) Behavior of the respondent unit

[17] See also OECD (2003, pp.28-30).
[18] For data collection techniques see chapter 6.

b) The instrument used to get information (kind of questionnaire)

c) The effect to the interviewer and, in general, the kind of survey technique.

All kinds of measurement errors are also called response errors (even though response errors should be referred to the first kind only).

7.19 Potential problems concerning the kind of unit can be summarized as:

a) Lack of capability to report correctly.

Errors due to the questionnaire may be caused by:

b) ambiguous wording of questions

c) Unclear layout of questionnaire.

Errors due to the respondent may be caused by:

d) insufficient knowledge to answer correctly

e) Lack of motivation to report correctly.

7.20 As regards the issue a), in the frame of BTS the business register from which the sample is drawn will usually consist of enterprises. Enterprises are therefore the usual *sampling un*its for BTS, i.e. the units that are selected for the sample. If the manager of the enterprise is able to complete the questionnaire for all the local units belonging to the enterprise, then the expected quality of responses will be high. Otherwise, it may be necessary to create a list of local units to whom the questionnaires should be addressed, provided that each local unit manager will be able to respond to the tendency survey questions.

In the first case (reporting units are enterprises); it may be difficult to provide estimates at a regional/local level. In the second case, sample local units data should be grossed up to the reference universe, which is normally based on enterprises rather than local units. Even though the recourse to kind of activity units may improve reliability of responses – since they correspond to specific production processes – in practice it is difficult to identify and use them as reporting units in the frame of a current survey. The main consequence is that the current sample may be a mixture of enterprise and local units, while sampling weights used for the grossing up procedure may have been defined at the enterprise level only. Moreover, if the sampling reporting unit list is not updated frequently, the risk to include in the sample too many local units of the same enterprises will be high.

As regards CTS, a parallel problem is due to frequent need to select the sample from the household universe rather than the consumers one. It may happen that the person belonging to the households selected for responding is not completely aware of the household economic situation. Moreover, the

sampling weights of the original sampling design are applied to household rather than to consumers, so that some biases may occur in the grossing up procedure.

7.21 Problems b) and c) can be reduced to a large extent by using questionnaires which have been previously tested and harmonized in order to avoid misunderstandings and wrong responses.

7.22 Problem d) can be minimized by using the kind of activity unit (KAU) as the reporting unit and collecting information from members of top management or their close associates. Users may sometimes demand information which, though useful for analytic purposes, may go beyond what the respondent can reasonably be expected to know.

7.23 The best way to minimize cause (e) is to demonstrate that the data produced from BTS are useful for the enterprises themselves. To do this, organizations conducting the surveys need to identify what information enterprises want to obtain for their own purposes.

7.24 If the questions are designed carefully there is little risk of any serious measurement error in BTS. The reason for this is that most of the questions relate to an assessment of levels ("too high", etc.), or the direction of change ("up" etc.) and this information is less subject to error compared with data on levels or changes in quantitative terms.

7.25 It is not easy to detect the presence of measurement errors. This problem depends on the particular field of interest and the amount of available time and resources. A first way simply consists in comparing observed and true values (of course, when available). A second method consists in a replication of interviews using more expert interviewers. But also in this case we cannot exclude the presence of measurement errors also in the second survey wave. Finally, it should be stressed that the most efficient strategy for reducing measurement errors is to eliminate possible causes of such errors during the survey design stage. In this respect it is important to pilot test questionnaires, instructions to respondents, and processing procedures before starting the survey.

*Processing Errors*

7.26 Processing errors may be introduced during:

- data entry
- data editing
- data tabulation.

Methods for avoiding errors at the data entry stage depend on the data collection method used. For personal and telephone interviews, the best approach is to use CAPI and CATI techniques and to build logical and consistency controls into these systems. When respondents are asked to report directly to a computer from a telephone, the same types of controls should be built into the computerized dialogue.

Regardless of the method of data collection used, the questionnaire should be designed so that correct data entry is facilitated. This means that questions and replies should be close together without any ambiguity as to which question a particular reply refers to. It is strongly recommended that practical tests involving the persons actually performing the data entry should be made before the questionnaire is finalized.

7.27 In the context of any data editing process, the main risk is that errors may be introduced by making the wrong adjustment. The risk of introducing errors into the data, by changing correct data at the editing stage, is best avoided by strictly adhering to logical controls and checking apparent logical errors and other inconsistencies with the respondents before any adjustment are made. The general rule is that, whenever possible, editing should be done at the same time as the data are entered in the database. In this way, errors at registration can be detected and eliminated directly when they arise.

It is important to record the errors that are detected in the editing process. This helps to identify areas where improvements are required in the questionnaires or in the instructions to respondents. In general, the need for editing of tendency survey information is significantly less than that required for quantitative surveys, because it is easier for respondents to supply the information that is being sought.

7.28 The risk of error at the data tabulation stage arises due to use of incorrect estimation criteria, or incorrect programs for processing the individual records. The first cause can be avoided by using the correct methodology and by working together with IT specialists developing the data processing software for the survey. The second problem is avoided by testing the software systems on a trial set of data before accepting it for actual use.

7.29 Efficient micro-editing of tick-box data cannot be achieved with significance editing. For quantitative data, the large range in values of the errors from responding businesses means a small proportion of the units tend to be responsible for most of the error generated in the statistics. It is desirable to concentrate most editing effort on these *significant* units' responses. This situation does not occur for tick-box values because they can only be 0 or 1. Any error in a tick-box response is

roughly as significant as an error in the next. It is not possible to obtain large gains in accuracy by editing only a small proportion of tick-box responses. However, this assumption fails if the estimation procedure adopted is based on a weighting system, according to which each responding unit receives a weight assessing its importance on the overall sample. For instance, in BTS weights can be given by the number of persons employed, on the basis of the assumption that the responses provided by the contact person can be extended to all the other employees of the firm. In this case, large gains in accuracy may be obtained giving priority to large firms.

Broadly speaking, automatic editing should be used in order to correct tick-box edit failures. The technique involves using algorithms to find the least number of data values (from those that failed the edits) which, when corrected, allow the complete set of failed data to pass the edits. The data items requiring correction are replaced by imputed values. Therefore, a tool for editing tick-box data must be able to select a minimum set of failed data requiring correction and create the imputed values required.

**(iii) Non-responses**

7.30 Non-responses are due to the refusal, or the not possibility to respond by a subset of sample units. Non-responses have two main consequences: 1) increase of the sampling error, since the estimate variance increases if the number of respondent units decreases; 2) effects on the non-sampling error, due to the potential bias derived from the fact that the average profiles of respondent and not respondent units are different.

7.31 To reduce non-sampling and sampling errors, a limitation of non-response should be achieved. In addition the survey set-up should be designed in such a way that non-response of a specific group of respondents with different response behavior (large deviation of answers to the responses from the other respondents) is avoided. Non-response can be divided in two categories:

*Item non-response* (partial non-response or missing value): sometimes respondents don't complete the whole survey because they are not willing or able to answer a given question.

*Unit non-response* (total non-response or missing record): respondents are not able or willing to cooperate with the survey.

7.32 Preventing unit non-response is a crucial element of data collection. During the fieldwork period, a constant monitoring of the response rates is necessary. When the response rates are low, the data collection strategy foresees a strategy to improve response rates, deals with large firms,

give rules for re-interviews and eventually non-responding firms or consumers can be substituted. Chapter six deals with these elements of data collection.

7.33 Even if in some countries answering to surveys conducted by public institution is mandatory by law, it is not advisable to force person to answer as the quality of the results strongly depends on the willingness to individuals to participate. It is advisable, instead, to make efforts in motivating and convincing participants of the usefulness of their contribution, also by letter and /or with small remunerations.

7.34 However, unit and item non-response will be present when the data collection period is finished and the data is being processed. There are a number of ways to deal with missing values. One of these is to *impute* a valid value for the missing value in the data file, while an alternative is to leave these values unknown. Reasons to impute a value, instead of leaving the field empty, are:

a) To obtain a 'complete' (completely filled) data file. Obtaining a complete file, with complete records, makes aggregation and tabulation easier, and prevents inconsistencies when tabulating.

b) To increase the quality of the micro file and/or of the parameter estimates.

7.35 If we want to use imputation to improve the quality, 'the quality of what' should be clear. Often, the primary goal is to accurately determine means and totals. We may also want to determine the distribution of a variable. For academic research purposes, it is important to have a good micro file, on which researchers can perform a variety of analyses. Different objectives can lead to different 'optimum' imputations.

7.36 Response rates are useful tools to monitor the data collection process and are quality indicators for the survey. The response rates, the ratio between answered and formulated questions, can be measured in different ways, each with its own purpose. Units which do not respond because they have ceased to belong to the target universe (terminated, switched to another (main) kind of activity, etc.) are not part of non-response (OECD, 2003, par.76). The simplest measure of non-response (*NR*1) is the percentage of enterprises in the sample from which information for the actual survey were not obtained. It is defined as:

$$NR1 = \left(\frac{n\prime}{n}\right) \cdot 100 \qquad (5)$$

where: $n$ = the number of enterprises in the survey; and

$n$' = the number of enterprises which did not submit usable information.

7.37 (OECD, 2003, par.77) This measure is useful for checking the efficiency of the data collection procedure. It is also a good indicator of the importance of non-response in censuses and sample surveys with uniform sampling fractions where all reporting units have equal weight.

7.38(OECD, 2003, par.78) For sample surveys with different inclusion probabilities for different enterprises, and for surveys where answers are weighted according to the size of the reporting units, measure *NR*1 is *not* a good indicator of the importance of non-response. For sample surveys with unequal inclusion probabilities for different units but equal weights for all units, a proper measure of non-response (labelled *NR*2 below) is then calculated as follows:

$$NR2 = \left(\frac{\sum_{i=1}^{n'} \frac{1}{f_i}}{\sum_{i=1}^{n} \frac{1}{f_i}}\right) \cdot 100 \qquad (6)$$

where: $f_i$ = the sampling probability for unit $i$ .

7.39(OECD, 2003, par.79) If, in addition, responses are weighted by the size of the reporting units, the correct measure to use (here called *NR*3) is:

$$NR3 = \left(\frac{\sum_{i=1}^{n'} \frac{1}{f_i} w_i}{\sum_{i=1}^{n} \frac{1}{f_i} w_i}\right) \cdot 100 \qquad (7)$$

where: $w_i$ = the size weight for unit $i$ .

7.40 (OECD, 2003, par.239) The OECD manual states that, assuming that the initial panel is selected as a stratified random sample and is then updated at regular intervals, the minimum response rate should be at least 50%. Without the use of a fixed panel, however, the response rate will need to be somewhat higher – 60 or 70%.

7.41 Often the combination of *NR*1 with *NR*2 or *NR*3 can already provide good indication on the distribution of the collected responses across company weights. *I.e.*, when the un-weighted non-response rate (*NR*1) is lower than the weighted non-response rate (*NR*2 or *NR*3) one or more of the larger enterprises have not completed their questionnaires yet. Then, targeted actions to collect the answers to the questionnaires of these larger enterprises can be undertaken.

**B. Treatment of non-responses**

**(i) Total non-responses**

7.42 The main problem is the possible bias due to non-respondents. It can be faced is a way similar to that seen for coverage errors[19]. We indicate as $P_R$ the relative frequency of respondent units (and as 1-$P_R$ the relative frequency of those not respondent); if $\bar{y}_R$ and $\bar{y}_{NR}$ are, respectively, the mean of $y$ in the 2 parts of the population, we will have:

$$\bar{y} = \bar{y}_R P_R + \bar{y}_{NR}(1 - P_R) \tag{8}$$

Similarly to what seen in subparagraph A (i), if we estimate $\bar{y}$ using the estimator $\hat{\bar{y}}_R$ which is an unbiased estimator of the subpopulation mean $\bar{y}_R$, based on the respondent units only, the bias of the estimator $\hat{\bar{y}}_R$ is given by:

$$E(\hat{\bar{y}}_R) - \bar{y} = (1 - P_R)(\bar{y}_R - \bar{y}_{NR}) \tag{9}$$

7.43 It depends on the share of units not respondent (in the population) and the difference between the means in the two subpopulations. The bias is not dependent on the sample size. A lower effective sample size leads to a higher variance. Since variance could become small for large samples, probably in large samples the relative weight of bias on the global error could be higher respect to that of sampling variance. In practice, we can evaluate bias only if we can use sources external to the current survey.

7.44 A first tool for tackling non-responses simply consists in increasing the number of responses using call backs. But call backs have a cost and they must be carried out within a limited time horizon. In practice, one often substitutes non respondents with other units chosen at random.

7.45 A second well known methodology is re-weighting of respondents. Under the assumption that propensity to respond is related to levels of the variable object of interest, we suppose to split the population in $L$ strata, in order to work with post-strata with quite different rates of response and average levels of the target variable inside. Strata are defined as *adjustment cells* from non-response. The post-stratified estimator will be:

$$\hat{\bar{y}}_{ps} = \sum_{h=1}^{L} W_h \hat{\bar{y}}_{Rh} \tag{10}$$

where $\hat{\bar{y}}_{Rh}$ is an unbiased estimator of the respondents' mean in the population post-stratum $h$. Bias is given by:

[19] In the literature it is distinguished between deterministic and stochastic non-response. For the discussion of these concepts we refer to the literature in the references of this chapter.

$$B(\hat{\bar{y}}_{ps}) = \sum_{h=1}^{L} W_h P_{NR}(\hat{\bar{y}}_{Rh} - \hat{\bar{y}}_{NRh}) \qquad (11)$$

where $P_{NR}$ is the non-response rate in stratum *h*.

In order to minimize the above bias, we must have that: a) in the same stratum means of respondents and non-respondents are quite similar; b) the response rates are quite different among strata (if they are all equal the bias (11) turns out to be equal to (9), so that bias cannot be reduced).

7.46 Crucial aspects concern the choice of the variable(s) to be used for carrying out post-stratification (that, of course, must be available both for respondents and non-respondents) and the need to estimate weights *W*. Generally speaking, reweighting does not necessarily increase the variance of final estimates. Admittedly, when weights need to be estimated from the sample, this leads to additional variance. This effect may be counteracted if auxiliary variables used to define adjustment cells are sufficiently related to the survey outcome (Roderick, Vartivarian, 2005).

7.47 A third possibility is based on *re-weighting* of respondents. The estimator is:

$$\hat{\bar{y}}_{rw} = \frac{\sum_{i=1}^{n_R} y_i / \pi_{hi}\tau_{hi}}{\sum_{i=1}^{n_R} 1/\pi_{hi}\tau_{hi}} \qquad (12)$$

where $n_R$ is the number of respondents, $\tau_{hi}$ is the unit response probability - given that the *i-th* unit in the post-stratum *h* has been included in the theoretical sample with inclusion probability $\pi_{hi}$. The product $\pi_{hi}\tau_{hi}$ is the estimate of the *final response probability* (two phases sampling), so that (12) is a common Horvitz-Thompson estimator. Response probabilities can be estimated using *logit* or *probit* models: this task can be tackled especially when panel surveys are carried out.

**(ii) Partial non-responses**

7.48 While total non-responses are treated using re-weighting, partial non-responses are treated using imputation. The easiest way consist in deriving the missing value from information available in other parts of the questionnaire when it is feasible. In general, imputations are predictions for the missing values, based on a model. Here a few different imputation methods will be described.

7.49 The choice of the imputation method should be made in consideration of the type of question asked (categorical or quantitative) and depending on the available auxiliary information.

You can consider different auxiliary information in order to evaluate their efficiency for the imputation. For example, initially you can evaluate information about the same variable in a previous period (for the same record) and information from other sources; then knowledge about the same variable from the item respondents can be considered.

7.50 In *deductive* or *logical imputation* you examine whether it is possible, based on logical or mathematical relationships between the variables, to unambiguously derive the value of one or more of the missing variables from the values that were observed. This method can also be applied if the rule does not necessarily always holds true, but only very probably does so.

7.51 For deductive imputation, it is not necessary to specify or estimate models. With only the edit rules as input, the process can be performed completely automatically. Furthermore, deductive imputations are, in a way, the best possible imputations since they are exactly equal to the actual values if the other values in the record are correct. Given this last assumption, it is important to perform the method after as many errors as possible have been detected and then corrected (systematic errors), or designated as 'missing'.

7.52 Deductive imputation is then the most logical subsequent step. Model-based and donor methods can be used afterwards. For estimating the parameters, these methods can profit from the values already filled in deductively. In view of the advantages of the method, it will always have to be determined what options there are for deductive imputation.

7.53 For business and consumer tendency surveys, these method is only limited usable. In some cases, if the answering scheme allows for 'not applicable' or 'not known' then this can be used for imputations. However, for most questions, this is not the case.

7.54 In *mean imputation,* a missing value is replaced by the mean score on the variable concerned for objects that have a valid score. Mean imputation does lead to a peak in the distribution, because the same mean is imputed for each missing value. Thus, the imputed value $\tilde{y}_i$ for a missing score $y_i$ in mean imputation is equal to the observed mean:

$$\tilde{y}_i = \bar{y}_{obs} \equiv \frac{\sum_{k \in obs} y_k}{n_{obs}} \qquad (13)$$

where $y_k$ is the observed score of the $k^{\text{th}}$ respondent and $n_{obs}$ the number of item respondents for variable $y$.

7.55 If desired, the objects can be weighted unequally, for example, due to differences in the inclusion probability. The resulting imputation is then usually a better, less biased estimator of the population mean.

$$\tilde{y}_i = y_{obs}^{(w)} \equiv \frac{\sum_{k \in obs} w_k \cdot y_k}{\sum_{k \in obs} w_k} \qquad (14)$$

where $w_k$ is the weight of the $k^{\text{th}}$ respondent.

7.56 This method is only recommended if no auxiliary information is available or when the available auxiliary variables are only marginally associated with the imputation variable. If the fraction of missing values on a variable is very small, and the imputations will have a marginal effect on the parameter to be estimated (such as the population total). Mean imputation may be permissible due to efficiency considerations, however, using this rather overly simplistic method should be an exception.

7.57 In *group mean imputation*, a missing value is replaced by the mean score on the variable concerned for objects that have a valid score *and* are in the same subpopulation as the item non-respondent (or in the same stratum). Consequently in group mean imputation there are a number of smaller peaks.

$$\tilde{y}_{hi} = \bar{y}_{h;obs} \equiv \frac{\sum_{k \in h;obs} y_{hk}}{n_{h;obs}} \qquad (15)$$

where $y_{hk}$ is the observed score of the $k^{\text{th}}$ respondent in group $h$ and $n_{h;obs}$ the number of item respondents for variable $y$ in group $h$.

7.58 Auxiliary information is used in group mean imputation, which involves classification into groups (subpopulations, imputation classes) based on one or more qualitative variables. The more homogeneous the subpopulations with respect to the variable to be imputed, the better the imputations, given the assumption that the classification into subpopulations not only effectively discriminates among respondents, but also among non-respondents. In group mean imputation, the peak of the distribution is usually much smaller, since the variation between groups is included in the imputation; only the variation within groups is disregarded.

7.59 Assuming that BTS will generally use panels, *longitudinal imputation* models can also be used. Longitudinal imputation is distinct from the earlier described methods because, during the imputation, use is made of data from the same unit at different times, often without using data from other units.

7.60 There are two main reasons to use longitudinal imputation techniques instead of the cross-sectional methods:

a) earlier or later observations of the same object are very good predictors for the missing value. This means that the quality of the imputation can be strongly improved
b) generally we not only look at longitudinal data cross-sectionally, but we are also interested in changes over time. To correctly estimate these changes, it is important that the imputation takes into account previous and future values.

7.61 One point that must be taken into account for longitudinal data is the way in which new information must be dealt with. In a longitudinal data file, the best possible imputation at micro level is obtained if as much information as possible from the past (and the future) is included. If, therefore, new information comes in, such as a new wave of data for a panel, then this new information should be used to revise or improve the values already imputed.

7.62 LOCF *(Last observation carried forward*) is a method that is frequently applied because it is very easy to use. In this method, the last observed value of an individual is used for the values of all later periods that must be imputed.

$$\tilde{y}_i^t = y_i^{t-1} \qquad (16)$$

7.63 This method is mainly applicable to categorical variables for which it is known that they change very little or not at all over time. For other categorical and quantitative variables, this method often mistakenly produces an overly stable picture of the actual situation. For example, for index figures, this method can lead to the observation of a non-existent stability. For BTS, this method can be used for the quantitative variables like months' work assured or capacity utilization rate.

7.64 Ratio imputation is frequently used for longitudinal data for which it is reasonable to assume that the observation at period *t* is proportional to the observation at period *t-1*. This method, which is frequently used in economic statistics, can be considered a refinement of last observation carried forward, in which corrections are also made for general changes over time.

$$\tilde{y}_i^t = \frac{\sum_{obs} y_i^t}{\sum_{obs} y_i^{t-1}} \cdot y_i^{t-1} \qquad (17)$$

7.65 Just as in mean imputation, ratio imputation can be applied separately per subpopulation (imputation class). This is done mainly if the ratios between the subpopulations vary strongly. It

should also be noted that a different disturbance term for each time period can be selected every time. Ratio imputation can be applied for missing values on a quantitative variable.

7.66 Finally, methods recommended by DG ECFIN and OECD for the treatment of missing data in ETS are provided hereinafter. DG ECFIN recommendations are:

- Institutes should describe in their metadata the precise nature of the procedures used in the treatment of item non-response or missing data.
- As a minimum requirement it is recommended that institutes closely monitor the impact of missing data (especially for large firms) and develop a clear set of strategies to minimise non-response.
- The use of imputation methods for the treatment of remaining missing data should be considered with care, in order to avoid possible distortions.
- Re-weighting techniques, taking account of different composition of the panel in adjacent surveys, are recommended as a means of reducing bias.

7.67 For qualitative questions, the recommended method by OECD is to assume the same distribution over the response alternatives [(+), (=) and (–)] as the responding report units in that industry. For questions requiring answers in percentages or numbers, assume that the non–responding report units have the mean value of responding report units in that industry.

## References

Cicchitelli, G., Herzel, A., Montanari, G.E. (1992). *Il campionamento statistico*. Il Mulino, Bologna.

De Waal, T., Pannekoek, J., Scholtus, S. (2011). *Handbook of Statistical Data Editing and Imputation*. J.Wiley & Sons, New York.

OECD (2003). *Business Tendency Surveys: A handbook.* OECD Publications, Paris, ISBN 92-64-19894-6. (http://www.oecd.org/std/leading-indicators/31837055.pdf.

Roderick, J. L., Vartivarian, S. (2005). *"Does Weighting for Nonresponse Increase the Variance of Survey Means?"*, in *Survey Methodology*. Catalogue No. 12-001-XIE, Statistics Canada, December.

Särndal, C.E., Lundström, S., (2005). *Estimation in Surveys with Nonresponse*. J.Wiley & Sons, New York.

# CHAPTER 8

# Processing tendency survey data

## A. Conversion of multiple choice questions into a time series

8.1 ETS are qualitative instruments in the sense that respondents are generally[20] asked to assign "qualities" rather than quantities or precise figures in response to given questions. In this respect, the answer scheme provided for qualitative questions is a multiple choice scheme. More specifically, for BTS, three possible answering options are usually provided: positive (increase, more than sufficient, etc.), neutral (remain unchanged, sufficient, etc.) or negative (decreased, not sufficient, etc.). For CTS five/six possible answering options are provided: the three options used for BTS plus "very positive" (got/get a lot better, very much higher, increase sharply, etc.), "very negative" (got/get a lot worse, very unfavourable, fall sharply, etc.) and "don't know" (for further details, see chapter 3).

8.2 A relevant aspect in this type of surveys is the quantification of the answers. Consumers and business tendency surveys asking qualitative/ordinal-response questions provide important information about respondents' expectations and perceptions on many aspects of overall business and economic conditions. Quantified information extracted from these surveys, e.g., the University of Michigan's Consumer Sentiment Index and consumer inflation expectations, are often used by researchers and policy makers in forecasting macroeconomic aggregates (Breitung, Schmeling, 2013) and testing canonical economic theories like the rational expectations hypothesis (REH) (Lahiri, K., Zhao, Y., 2013). Since the seminal paper of Theil (1952), researchers have long been scrutinizing various quantification methods: a review of the most widely used quantification methods can be found in Nardo (2003), which focuses on the shortcomings of different approaches to modelling ordinal responses and in Pesaran and Weale (2006). In the next sections, a general exposition and review of quantification methods is provided.

[20] In the business and consumer tendency surveys, some quantitative questions are requested, too. For details see Chapter 3.

**(i) The balance statistic**

8.3 A very common and widely used method for the conversion of multiple choice questions into a time series is the calculation of the aggregate balance for each question. Balances (EC, 2014) are the difference between positive and negative answering options, measured as percentage points of total answers. In particular, if a question has three different options, "positive", "neutral" and "negative" and if P, U and N (P+U+N=100) denote the percentages of respondents having chosen respectively the option positive, neutral and negative, the balance is calculated as follows:

$$B=P-N$$

8.4 In the case of questions with six options (CTS), the balances are calculated on the basis of weighted averages. If P, U and N have the same meaning described above, and PP denotes the percentage of respondents having chosen the option "very positive", NN the percentage of respondents having chosen the option "very negative" and NK is the percentage of respondents without any opinion (PP+P+E+N+NN+NK=100), balances can be calculated as[21]:

$$B=(PP+1/2*P)-(1/2 * N + NN)$$

8.5 Balances can vary from -100, when all respondents choose the negative option (or the most negative one in case of six options questions) to +100, when all respondents choose the positive (or the most positive) option.

8.6 Business and consumer tendency survey data are usually analysed on an aggregate basis, considering the "balance" calculated as described above. The balance provides an easy-to-compute, easy-to-understand quantification of survey results, and as such is considered as a useful tool to analyse business cycles.

**(ii) Diffusion indexes**

8.7 Another measure commonly used is the diffusion index, which is another way of quantifying the information contained in the survey. Generally speaking, diffusion indexes measure the proportion of the components that contribute positively or negatively to the index. Referring to tendency surveys, a diffusion index indicates the degree to which the indicated change is diffused throughout the sample.

[21] This is the formula commonly used at European level. Other weights can be used, too. E.g., Italy uses the following formula: B=(2*PP+P)-(N+2*NN). In this case, the balance varies between -200 and +200.

8.8 There are different ways to calculate a diffusion index using ETS data. The most common consists of taking the percentage of respondents reporting a "positive" answer and adding it to one-half of the percentage of respondents reporting "unchanged". In this case, the index can vary from 0 to +100, the midpoint being 50. The formula that links the balance to the index is the following:

$$B=2*(DI-50) \text{ and } DI=(100+B)/2$$

where B is the balance and DI is the diffusion index[22].
Both indicators move in the same direction over time (one is just a linear transformation of the other) but their ranges are different and hence diffusion indexes are usually flatter than balances when shown graphically.

8.9 The Purchasing Managers' Index (calculated by Markit Group in the key economic sectors of many countries) is an example of composite indicator calculated by diffusion indexes. For example, for the United Kingdom (UK) manufacturing sector, Markit UK calculates a PMI composite indicator with the following diffusion indexes: production level, new orders, supplier deliveries time, stocks of items purchased, employment level, with the deliveries time index inverted so that it moves in a comparable direction; the weights of the indexes are 0.25, 0.30, 0.15, 0.10, 0.20 respectively. A PMI reading above 50 indicates an overall increasing in the economic activity, below 50 indicates an overall decrease.

**(iii) The probabilistic approach and other methods**

8.10 Two of the main shortcomings of the balance statistics are some strict assumptions about answers distribution: the "stay the same" U share is constant over time and the relationship among P (positive) and N (negative) is linear and constant over time. For this reason, various alternatives have been discussed in the literature, including the Carlson-Parkin probabilistic approach (Carlson and Parkin, 1975) the Pesaran (1984) regression approach, and the latent factor approach (D'Elia, 2005). However, it has to be noted that these methods can lead to measurement errors. The fact that they are linked to a reference series (generally the value to which the respondents should refer when they express an assessment) implies that these quantification methods can become unreliable when

[22] The conversion formula derives from: B=100*(P-N) and DI=100*(P+U/2); P+U+N=100, where P is Positive reply, U is Unchanged reply and N is the Negative one.

exceptional events impact heavily on the correlation between the survey data and the reference 'hard' data[23].

8.11 *The probabilistic method*. The probabilistic approach has been first proposed by Carlson-Parkin (1975) for survey questions with three possible reply options ($J = 3$), e.g., "up", "same", "down", or "good", "normal", "bad", without (utilizing) individual/consumers-level data $Y_{it}$. In absence of individual specific information, Carlson-Parkin assume some simplified hypotheses that can be summarized as follows (Breitung and Schmeling, 2013):

a) opinions of respondent i at period t are independently and identically distributed as $y_{it} \sim N(\mu_t;\sigma_t^2)$

b) the respondents report $r_{it} = 1$ (increase) if $y_{it} > \delta^+$, $r_{it} = -1$ (decrease) if $y_{it} < \delta^-$ and $r_{it} = 0$ (no change) if $\delta^- \leq y_{it} \leq \delta^+$

c) the number of respondents $N_t$ in period t is sufficiently large such that $n^+_t / N_t \approx p^+_t$ and $n^-_t / N_t \approx p^-_t$, where $n_t^+$ ($n_t^-$) denotes the number of respondents in t reporting an increase (decrease), $p^+_t = \text{Prob}(y_{it} > \delta^+)$ and $p_t^- = \text{Prob}(y_{it} < \delta^-)$. Notice that $\delta$ is constant over time.

Let $\Phi(z)$ denote the cumulative distribution function of the standard normal distribution and define the inverse normal scores as:

$$q_t^+ = \Phi^{-1}(1 - p_t^+) = (\delta^+ - \mu_t) / \sigma_t \quad (1)$$

$$q_t^- = \Phi^{-1}(p_t^-) = (\delta^- - \mu_t) / \sigma_t \quad (2)$$

Solving for $\sigma_t$ yields:

$$\sigma_t = (\delta^+ - \delta^-) / q_t^- - q_t^+ \quad (3)$$

and replacing $\sigma_t$ in equation (1 ) or ( 2) we obtain:

$$\mu_t = \delta^+ z_{1t} - \delta^- z_{2t} \quad (4)$$

where $z_{1t} = q_t^- / (q_t^- - q_t^+)$ and $z_{2t} = q_t^+ / (q_t^- - q_t^+)$.

[23] Friz R. and S. Lindén (mimeo), , "Can quantification methods lead to wrong conclusions? Evidence from consumers' inflation perceptions and expectations", preliminary work presented at the FRBNY Conference, November 2010, available at:http://www.ny.frb.org/research/conference/2010/consumer/FRBNY_Friz_Linden_draft_1Nov2010.pdf.

The average $\mu_t$ represents the value to which all the respondents should refer when they express an assessment on variable X. It is a linear combination of values deriving from a transformation of the observed frequency of the answers[24].

8.12 *The regression method*. The basic idea is to assume that operators implicitly attach to each qualitative answer also a numeric value of the variable (D'Elia, 2005). For instance, they reply "It did (will) decrease" when x decreases by 5%, and "It did (will) increase" when it grows by 10%. Of course, reference values of x corresponding to each possible qualitative answer are seldom available. However, when a traditional quantitative estimate of x is available, it is possible, in principle, to regress the latter indicator against the time series of the percentages of various types of answers. Then, the parameter of each explanatory variable can be regarded as a statistical estimate of the unknown reference level of x associated to this type of answer. Of course, timely quantitative measures of x are not strictly requested, since only a reliable sample of data is needed in order to run the regression. In addition, the regression model may be augmented by a list of other explanatory variables, such as a time trend, seasonal dummies, etc..

8.13 The regression approach has been introduced by Anderson (1952), and popularized by Pesaran (1984, 1987). Let assume that $P_t = (p_{1t}, \ldots, p_{kt})$ is the vector of the percentages of each type of answers given to the qualitative survey carried out at time t, and $Y_t$ is a standard quantitative measure of the opinions (or expectations) $x_t$, then the regression model

$$Y_t = a_1 p_{1t} + \ldots + a_k p_{kt} + u_t \qquad (5)$$

where $u_t$ is a random disturbance, provides a statistical estimate of the parameters $a_j$[25] . Of course, the estimates of $a_j$ are complicated functions of the observations on $P_t$ and $Y_t$. For instance, in the simplest case of ordinary least squares estimator, $\hat{a} = (\hat{a}_1, \ldots, \hat{a}_k)$ is given by the formula $\hat{a} = \Sigma^{-1}\Phi$, where $\Sigma$ and $\Phi$ are the variance matrix of the time series $P_t$ and the covariance between $P_t$ and $Y_t$ respectively. Of course, the percentages $P_t$ may enter the regression model (5) also after some algebraic manipulation, such as logarithmic, logistic or exponential transformation. In this case, the parameters â cannot be interpreted as the average level of x attached by the interviewed persons to each qualitative answer.

[24] Various modifications of this setup have been proposed in the literature; see again Nardo, 2003, for a survey of the literature.

[25] The constant term is dropped from the regression model, since $p_{1t} + \ldots + p_{kt} = 1$ by definition.

8.14 The regression approach is very simple and is based on a smaller amount of assumptions compared to the probabilistic approach. Furthermore, it allows treating the percentage of non-response as any other time series of percentages, in a way which is internally consistent. However, the regression approach requires the availability of some quantitative measure of the relevant variable, and this requirement could be very restrictive, especially in the case of expectations or "business climate". In addition, the components of $P_t$ are generally highly correlated to each other, thus multicollinearity is very likely to flaw the estimates of parameters. Sometimes, sequential differentiation of time series and some algebraic transformations may reduce this problem.

8.15 *The latent factor approach*. This method regards the percentages of each qualitative answer as a function of a common "latent measure" of x observed by people but not by statisticians. Usual multivariate techniques may help in estimating the dynamics or the sectorial variations (but not the absolute level) of the latent factor affecting the opinions and expectations expressed by the interviewed operators. As matter of fact, the time series of percentages of answers collected in qualitative surveys are highly correlated. First of all, this fact implies that the latent variable approach is possibly sound and reliable. However it also suggests that even very sophisticated methods, based on complicated transformations of original percentages, tend to produce indicators that, in fact, follow the common trend and cycle that can be easily deduced by whatever time series of percentage, or a simple combination thereof. This fact explains and justifies the widespread use of the "balance" between the percentages of "optimistic" and "pessimistic" answers[26].
For further details about the methods described, it refers to specific literature mentioned above in the text.

**(iv) Latest proposals for the quantification of qualitative information**

8.16 A completely different approach with respect to the methods described above, the method based on the reply patterns of individual firms was proposed in a recent study by Crosilla-Malgarini (2011). Balances do not fully exploit the information content of the survey, failing to properly consider "neutral" answers nor giving any attention to the possible relationship among different answers to the various questions at the firm level. Indeed, a more careful study of this information can provide interesting insight into firms' opinions.

8.17 More specifically, cluster analysis is used in order to identify various "behavioural models" according to which individual responses may be classified. The identification of the models, and the study of their evolution over time, has allowed an analysis of survey data that may well complement

[26] For an empirical evidence on the performance of the methods described above see D'Elia, E.(2005).

the more usual one based on the calculation of the balances. The method uses individual responses to a selected number of qualitative questions contained in the questionnaire of the manufacturing survey monthly carried out by ISTAT[27].

8.18 Considering six variables (questions) selected for the analysis (six questions: assessments on the total, internal, foreign orders and on production; expectations for three months ahead on production and orders), ISTAT chose to concentrate on the most economically significant behavioural models possibly emerging from the data (see the following table)[28]: this choice allows an easier interpretation of the results, allowing also the comparison of models over time. The nine models are represented by the nine centres (or representative objects) imposed in the clustering procedure[29]. In the table T8.1 shown below, the nine centres are described.

**Table T8.1**

**Behavioural models**

| | | **Expectations** | | |
|---|---|---|---|---|
| | | Positive | Unchanged | Negative |
| **Assessments** | Positive | PP | PU | PN |
| | Unchanged | UP | UU | UN |
| | Negative | NP | NU | NN |

8.19 The interpretation of the table is straightforward: the model PP represents a very positive stance expressed by the firms for both the current situation and that expected for three months ahead: after the application of the algorithm, all firms with positive opinions on the current and expected situation are included in this model (group); at the opposite side of the spectrum, when model NN do prevails, firms are expressing negative assessments and expectations on the six variables on which the analysis is performed. All the other models represent intermediate situations

[27] The specific application to the manufacturing survey is not a loss of generality for the method.

[28] The number of behavioural models is 729. In fact, the number of all the possible permutation of three elements in 6 questions is $3^6$. However, as a logical consequence of the results obtained by correlation analysis between selected variables, ISTAT considers only the major nine common association paths shown in the table T8.1.

[29] Cluster analysis groups the data according to the minimum distance between them, in other words, it looks for data associations by minimizing the "within group" variance and maximizing the "between groups" variance. In this study, the k-medoids method, a non-hierarchical clustering method, is used. This method is applied by PAM clustering algorithm (Partition Around Medoids carried out by software R) pre-imposing the "centres" (behavioural models or medoids) of the groups: the algorithm gathers the micro data minimizing the dissimilarity of the units of the dataset to the nearest pre-imposed medoid. A cluster is then defined as a subset of the original micro data assigned to the same medoid.

(model UP represent firms that answered "Unchanged" for all the assessments and "Positive" for all the expectations).

8.20 Finally, for a proper interpretation of results, the single units clustered together are weighted by the relative importance of the firm in terms of the number of its employees, in order to take into account the economic significance of each behavioural model.

8.21 Other alternative methods of quantification were proposed lately by Proietti and Frale (2010). More specifically, two new methods for the quantification of BTS concerning the qualitative assessment of the state of the economy are described. The first is a nonparametric method based on the notion of the spectral envelope; it originates a signal extraction filter which has solely cross-sectional dimension, i.e. only contemporaneous values are employed.
The second one is based on a dynamic cumulative logit model for the time series of ordered proportions. A cumulative logit unobserved components models featuring a common cycle are fitted by maximum likelihood. The conditional mean of the cycle, which can be evaluated by importance sampling, offers the required quantification.
For the application of the methods and for further details, it refers to specific literature mentioned above.

**(v) The "disaggregate" approach**

8.22 The method derives from the analysis of the panel dataset of individual (firm level) responses underlying the aggregate responses. It was originally proposed by Mitchell et al. (2002, 2004) and aims to produce an early indication of official output data, based on the fact that survey data are published ahead of official data on output growth. They found that more accurate indicators, so-called "disaggregate" indicators, can be obtained when quantification proceeds in a manner which allows for a degree of heterogeneity across firms.

8.23 The method includes two steps. First, the time series of individual (respondent – level) categorical responses are converted into quantitative series for the macroeconomic variable of interest, e.g. economy – or region - wide output growth. Second, the disaggregate indicator of the macroeconomic variable is derived by averaging these quantitative series across respondents at a given point in time. The expected value for the macroeconomic variable given an individual's categorical response is calculated in two ways. In the first mode, the Probability Density Function (PDF) of the macroeconomic variable conditioned on an individual's response is obtained by numerical integration; the expected value is derived by the conditional PDF. To calculate this

conditional expectation, under the parametric assumption that the PDF governs the target variable, the estimated relationship between the respondent's survey responses and the macro variable, using ordered discrete choice models at the firm level is used.

8.24 An alternative, non-parametric, way is to evaluate the conditional probability of output growth given an individual's survey response by taking the mean of the empirical distribution function. This is an intuitive approach: take the average (across time) of the macroeconomic variable when the individual replied "up", "the same" and "down". The quantitative series for this individual involves replacing an "up" with the average when the individual replied "up" and so on for "the same" and "down". The disaggregate indicator of the official macroeconomic variable is then defined as the average across respondents (either weighted or not weighted) of their expected values for the macroeconomic variable at a given point in time. For further details see EC (2006) and Mitchell et al. (2002,2004).

**B. The seasonal adjustment**

**(i) Seasonal patterns in opinion survey data**

8.25 *The need to adjust for seasonal influences*. As with many economic data, the variables investigated in business and consumer tendency surveys are subject to seasonal patterns. Seasonal patterns are upward and downward movements in time series that occur annually on a regular basis. Typically seasonal swings will obscure developments from one period to another of economical, labour market and financial developments. There is thus a need to remove these types of influences and this for two main purposes. First to see whether opinions of producers and consumers improve, stagger or deteriorate cyclically in the short run, i.e. from month to month or quarter to quarter. Secondly its only after seasonally adjustment that major turning points in the business cycle can be detected properly and also more easily.

8.26 However seasonal patterns might vary depending on the kind of economic activity and or on the kind of indicator. *Business managers* are aware of the seasonal fluctuations in their economic activity. Within the BTS, in many cases respondents are asked to take seasonal effects into account in their response, however not for all type of indicators regardless frequency, monthly or quarterly. For example in manufacturing industry managers are asked to do this with regard to the level of production, but for the assessment of their book of unfilled orders or stock of finished products they are not asked to do so. But even in cases that they are asked to take seasonal influences into account apparently respondents are unable to do this completely or sufficiently. For example

among a large part of the enterprises in manufacturing industry the level of production still shows seasonal patterns. Production is 'normally' lower in the first two months of the year and in the summer months compared to the remainder of the year.

8.27 Many outcomes of BTS variables exhibit seasonal patterns. Even though consumers are explicitly asked not to take into account such seasonal variations, in practice the answers frequently show seasonal patterns. Analysis of DG ECFIN of the EC on the BCS series in the harmonised programme indicated that approximately 45% of BCS series shows a seasonal pattern, 40% did not show a seasonal component and around 15% showed irregular behaviour (EC, 2006 P.60). The seasonal patterns in the data are usually weak and unlikely to change much over time.

8.28 *Consumers* are not asked to take seasonal influences into consideration when responding. The questions refer to developments in the past 12 months or expectations for the next 12 months, which (in theory) should automatically discard seasonal patterns. However, consumers' opinion data often also exhibit the swings which are 'normal' for the time of the year. Research in the field of psychology suggests that the 'mood' of consumers waver throughout the year.

8.29 *How to correct for seasonality?* For time series of quantitative monthly or quarterly statistics, the repeatedly seasonal patterns are adjusted by computing a year-on-year growth figure. This is one and often way used for seasonal adjustment. In this way, each month is compared with the corresponding month a year previously. Under the assumption that seasonal patterns are very similar or the same in each year this implicit method of seasonal adjustment gives a simple indication of the development of the indicator over time.

8.30 Although this year on year method is effective to remove the impact of the seasons, however on the short run it has some serious drawbacks, one of them being that for the short run the year information might be inadequate for analysis purposes. While on a year base the growth figures might still improve or deteriorate, the monthly or quarterly development can give a different picture, for instance in almost all cases the improvement or the deterioration is started earlier or that it already has developed in the opposite direction. In a year time a lot might have happen or changed and it is well known fact that only after explicit seasonally adjustment better, more actual developments can be traced. Furthermore, year-on-year comparisons are influenced by working and trading day effects and possible outliers. So, to adjust for this disturbing phenomena, again advanced adjustments needed to be done with software of seasonality.

8.31 A third and important drawback of this implicit method of seasonal adjustment, year on year growth, is that the figures are less reliable to determine turning points in the business cycle. A peak or through in the year growth rates simply indicates that a period of relative high or low growth has ended, nothing less nothing more. And with regardless to time series of balances this method is not recommended to do economic or cyclical analysis.

8.32 Another, more advanced or explicit way of seasonal adjustment, is decomposition of time series. Out of the data several main components are extracted such as the trend, the business cycle, irregularity and seasonality. All data collected in ETS should be systematically tested for seasonality and, if seasonality is detected, they should be adjusted using one of the standard seasonal adjustment software packages.

8.33 In the statistical community, seasonal adjustment is a much studied and debated issue. There are several different packages developed. The most common used packages are software based on X-12-regArima (ARIMA stands for *AutoRegressive Integrated Moving Average*) or TRAMO/SEATS (TRAMO stands for Time series Regression with ARIMA noise, Missing values and Outliers, SEATS stands for Signal Extraction in ARIMA Time Series). Dainties is used by the EC for seasonal adjustment of all BCS data in the harmonized programme. Other used software is STAMP (Structural Time Series Analyser, Modeller and Predictor) and software based on Structural Time Series Models (STM).

8.34 Seasonal adjustment is performed at the end of the statistical process. It is applied once the actual statistic production has been completed. For some statistics, both uncorrected and seasonally adjusted are being published. However, when possible, it could be beneficial to perform seasonal adjustment during the production process. In this way outliers can be detected and action towards respondents can be taken to improve the data collected.

8.35 As said the seasonally adjusted figures are easier to analyse a series over time and in the short run, because seasonal swings which can hinder this rather strong, are smoothed or filtered out of the raw data. For reasons of smoothness, an indirect seasonal adjustment should be done vis-à-vis a direct one. Suppose that the total manufacturing industry outcomes are calculated out of five underlying branches. First seasonal adjustment is only applied to the five underlying individual series, the adjusted outcomes are added up to get a result for the total series. This is called “indirect” seasonal adjustment. Next the original outcomes of the composite series are “directly” adjusted for seasonality not using seasonally adjusted data from any underlying series. If the indirect adjustment

gives a smoother result, as indicated by a measurement test for roughness or smoothness, then the directly adjusted series can be used for analysis purposes, else the outcomes of the direct adjustment will be preferred. In publications this must be made clear.

**(ii) General principles of seasonal adjustment of BCS data**

8.36 In general series that do not show seasonality should not be adjusted. The seasonal adjustment should be guided by the following general principles:

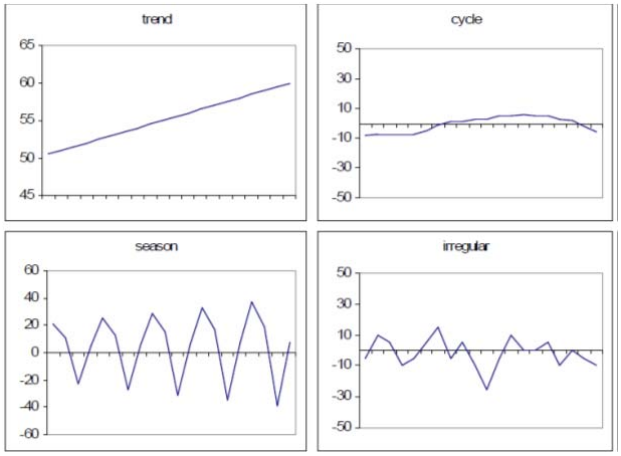
- seasonal adjustment should only be carried out if a pattern can be methodologically estimated with a reasonable level of accuracy, so that it can be removed effectively from the data
- no residual seasonal pattern should be left after the adjustment
- there should be no over smoothing
- it should not lead to abnormal revisions in the seasonal adjustment figure with respect to the characteristics of the series
- the adjustment process should prefer the simpler (ARIMA) models
- the underlying choices should be documented.

8.37 Seasonal adjustment involves the adjustment of a time series for influences that recur on an annual basis at fixed times with a certain intensity. To do this, we decompose a time series into its separate components, so that we can filter out the seasonal component. However, the separate components are not observable, because we only see the series in its totality. Seasonal adjustment is therefore subjective to a certain extent; there is more than one way of performing this decomposition.

8.38 It is assumed that a calendar and working day-adjusted time series (Y) can be broken down into the above four components: the trend component (T), the business cycle or cycle component (C), the seasonal component (S) and the irregular component (I).

**Figure F8.1**

**Four main components in time series**




**Source: Statistics Netherlands, 2011, Method series 11, Seasonal adjustment**

8.39 Various types of seasonal components can be distinguished in time series, not all of which are found in business and consumer tendency surveys data. Typically outcomes from tendency surveys don't possess the long term trend: balances waver in a band wide of +100 and -100 (except for a few variables such as utilisation degree or index order position). The trading day and holiday component is also not used in confidence surveys. The following components may be present in business and consumer tendency surveys:

a) The *incidental outlier* component, usually extracted prior to seasonal component. Outliers in BCS are limited by the boundaries in the answering categories (+/- 100 percent)
b) The *seasonal* component
c) The *business cycle*
d) The *irregular component* consists of fluctuations in the time series caused by coincidental non-systematic factors. This component contains the part of the data that is not part of one of the other components.

**(iii) Widely adopted solutions for seasonal adjustment**

8.40 There are a few main types of software for extraction of the seasonal component.

X12-RegARIMA developed by the US Census. X12-RegARIMA is actually based on a non-parametric approach of extraction of the seasonal component. The method is based on the iterative

estimation of the seasonal component of a time series by means of several calculation rounds. Using ARIMA models, the series is extrapolated to make better estimations, but the X12-procedure is based on empirical rules. Regression (Reg) is used for adjustment of calendar effects and other special influences.

8.41 TRAMO-SEATS developed by the Bank of Spain, a parametric approach and software package, in which SEATS is the module for seasonal adjustment. The method tries to model time series and estimate seasonal effects using ARIMA models. TRAMO makes use of regression for same reasons as X12, trading day and holidays and also several types of outlier adjustment. SEATS is used for seasonal adjustment and extracting the business cycle out of the trend-cycle.

8.42 A third approach for seasonality is Structural Time Series Models (STM). This approach is intended to extract the seasonal adjustment simultaneously with other components such as trend whereby each component in a time series can be modelled separately, and therefore the seasonal component is estimated separately. STAMP is a well-known software package that is used for this purpose.

8.43 Dainties is currently used by DG ECFIN for adjusting the short term economic statistics. In the Dainties procedure a seasonally adjusted series is obtained through a moving window of regression. A brief description of Dainties is in the Annex 1 of the ECFIN'S User Guide (EC, 2014).

8.44 National statistics institutes increasingly make use of either X12-RegARIMA or TRAMO-SEATS and in this they apply a set of common rules, guidelines and best practices. In practice, the same quality of seasonal adjustment can be attained using X12-RegARIMA or TRAMO-SEATS, at least for regular series without special effects. The Eurostat guidelines for example (European Statistical Office, 2009) therefore do not demonstrate a preference for either of the methods.

8.45 An advantage of the Dainties is that it is useful for data where the seasonality changes only very slowly or not at all. Furthermore Dainties does not lead to revisions, which is suitable as BCS data are not revised in general. A drawback is that no outlier correction is currently available for the Dainties method. Drawback of both X12 and TRAMO-SEATS is that they introduce revisions to historic seasonally adjusted series as they make use of forecasts of future observations which are later replaced by the actually observed values.

8.46 Research indicates that there are no marked differences in the performance of the seasonal adjustment methods applied to tendency surveys. Furthermore the data are close to "deterministic seasonality" which makes the Dainties method useful in this context. In total, X12 detected more seasonal series than TRAMO-SEATS in a DG ECFIN comparison. This was the case for all sectors except services.

## References

Anderson, O. (1952), *The Business Test of the IFO-Institute for Economic Research,Munich, and its Theoretical Model*, Review of the International Statistical Institute, 20, 1-17.

Breitung, J. (2008), *Assessing the Rationality of Survey Expectations: The Probability Approach*, Journal of Economics and Statistics (Jahrbuch fur Nationalökonomie und Statistik), 228, 6, 630-643.

Breitung, J., Schmeling, M. (2013), *Quantifying survey expectations: What's wrong with the probability approach?*, International Journal of Forecasting, 29, 142-154 Hannover, Germany, December.

Crosilla, L., Malgarini, M. (2011), *"Behavioural models for manufacturing firms: analysing survey data"*, in Journal of Industrial and Business Economic, vol. XXXVIII – n.4, 2011 December.

Carlson, J.A. and Parkin, M. (1975), *"Inflation expectations",* Economica, vol. 42.

D'Elia, E.(2005), *"Using the results of qualitative surveys in quantitative analysis"*, ISAE WP series n.56, September.

European Commission, Directorate-General for Economic and Financial Affairs (2006), *"The Joint Harmonised EU Programme of Business and Consumer Surveys"*, Special Report No.5.

European Commission, Directorate-General for Economic and Financial Affairs (2014), *"The Joint Harmonised EU Programme of Business and Consumer Surveys"*, user guide March, available at http://ec.europa.eu/economy_finance/db_indicators/surveys/method_guides/index_en.htm.

Eurostat (2009), *ESS guidelines on seasonal adjustment*.

Henzel, S.,Wollmershäuser, (2005) T., *An Alternative to the Carlson-Parkin Method for the Quantification of Qualitative Inflation Expectations: Evidence from the IFO World Economic Survey*, IFO Working Paper No. 9, June.

Ladiray, D., Quenneville, B. (2001), *Seasonal Adjustment with the X-11 method,* Springer-Verlag, New York.

Lahiri, K., Zhao, Y. (2013),*Quantifying Heterogeneous Survey Expectations: The Carlson-Parkin Method Revisited*, Department of Economics, University at Albany, SUNY, Albany, New York, 12222, USA, April.

Lothian, J. and Morry. M. (1978), *A set of Quality Control Statistics for the X-11-ARIMA Seasonal Adjustment Method*, Working Paper 78-10;

Mitchell, J., R.J. Smith and M. R. Weale (2002), *"Aggregate versus disaggregate survey-based indicators of economic activity"*, NIESR, Discussion Paper No. 194.

Mitchell, J., K. Mouratidis and M. R. Weale (2004), *"The impact of survey aggregation methods on the quality of business survey indicators"*, NIESR, London.

Nardo, M. (2003), *The Quantification of Qualitative Survey Data: A Critical Assessment*, Journal of Economic Surveys, 17(5), 645-668.

Organisation for Economic Co-operation and Development (2003), *Business Tendency Surveys: a handbook*, Statistics Directorate.

Pesaran, M. H. (1984), *Expectations formation and macroeconomic modelling, in Contemporary macroeconomic modelling*, P. Malgrange and P.A. Muet (eds.), Blackwell, Oxford.

Pesaran, M.H. (1987), *The Limits to Rational Expectations,* Oxford: Basil Blackwell.

Pesaran, M.H., Weale, M. (2006), *Survey expectations in handbook of Economic Forecasting*, North Holland, pp.715-776.

Proietti, T, Frale, C. (2011), *New Proposals for the quantification of qualitative survey data*, Journal of forecasting, Volume 30, Issue 4, pg.393-408, July.

Seitz, H. (1988), *The Estimation of Inflation Forecasts from Business Survey Data*, Applied Economics, 20(4), 427-438.

Statistics Netherlands, 2011, Method series 11, *Seasonal adjustment*, (http://www.cbs.nl/NR/rdonlyres/75B108E9-FE3E-4506-904D-166AD9F785A6/0/2011x37062art.pdf).

Theil, H. (1952), *On the time shape of economic micro variables and the Munich business test*, Revue de l'Institut International de Statistique, 20, 105–120.

U.S. Census Bureau (2007), *X-12-ARIMA Reference Manual, Version 0.3, section 5.5*, Time Series Staff Statistical Research Division, U.S. Bureau of the Census, Washington D.C. (http://www.census.gov/srd/www/x12a/x12down_pc.html).

# CHAPTER 9

## Data dissemination and publication

### A. Metadata

9.1 One of the best practices in the dissemination of data is the provision of metadata for the data produced or generated. By describing the contents and context of data files, the quality of the original data is greatly increased for users. The metadata provides information and explanations on how, by whom and for what purpose the data are produced or generated, the survey details, methodology, other details as well as possible limitations of the data. The metadata are very useful not only for the data producers or generators but definitely for all data users. A metadata set information should be provided with sufficient information about the overall characteristics of the survey to enable users, e.g., to draw accurate inferences from the data or to determine whether the results of surveys conducted in different countries are comparable. An efficient metadata set should be updated at least once a year.

9.2 For ETS, the main aspects that should be included in a metadata set information are listed in the following table[30].

**Table T9.1**
**ETS: an example of metadata set information**

| **Contact data** |
|---|
| Country |
| Survey |
| Organization |
| Official address |
| Website |
| Contact person |
| **Methodology** |
| Survey frame |
| Size of frame list |
| Characteristics |

[30] Most items of the table (above all those concerning the methodology) are in the metadata set supporting business and consumer surveys in the joint harmonized EU programme of the EC (for further details about metadata set requested by the EC see http://ec.europa.eu/economy_finance/db_indicators/surveys/metadata/index_en.htm).

| |
|---|
| Frame list update |
| Population |
| Sampling method |
| Sample size |
| Sampling error |
| Response rate |
| Treatment of non-responses |
| Sample coverage |
| Weighting scheme |
| Sectors and/or categories currently covered |
| Periodicity |
| Survey method |
| Fieldwork period |
| Remarks methodology |
| **Data and dissemination** |
| Unit of measure |
| Data revision |
| Frequency of dissemination |
| Dissemination format |
| Release policy |
| Metadata update |

**Source: European Commission, Eurostat**

The table is divided in three main parts: contact data, methodology, data and dissemination. In the following paragraphs, a detailed description of each item of the table is provided.

**(i) Contact data**

9.3 The contact data represent useful information on the national organisation conducting the survey such as the contacts person. In this section, the name of the organization that carried out the survey, its official address and its website are indicated. In addition, to facilitate contacts, in the item "contact person" all the data (name, unit organization, phone number, fax number, email address) of the person that may be contacted regarding the survey have to be indicated.

**(ii) The methodology**

9.4 In this part the information concerning the methodology is described. More specifically, data about population, frame list, sampling, accuracy and data collection are requested.

9.5 In *survey frame* (see chapter 4 for definition), you can specify the type of frame list used for the survey. E. g., for BTS: business register, government register, electoral rolls own sources of information etc. and official census, statistical registers, mobile phone register or fixed telephone lists, etc. for CTS.

9.6 In *size of frame list*, type (firm, activity unit, local unit, household, individual etc.) and number of statistical units of the frame list have to be indicated.

9.7 *Characteristics* that are considered in the frame list for each unit have to be indicated: addresses and contact details of enterprises, main kind of economic activity, number of employees, turnover etc. for BTS. For CTS: name, telephone number, age, gender, socio-economic group etc..

9.8 *Frame list update* means how often the frame list is updated: every year, every two years, 3-4 years etc.

9.9 In the item *population* the number of units in the target population have to be specified.

9.10 The selection method of units has to be pointed out in the item *sampling method*: probabilistic selection (representative sample, balanced sampling, random route sampling, systematic sample etc.), or deterministic ( purposive or judgemental methods, quota sample etc.) one. It is important describe sample design features: e.g., the presence of stages or stratification have to be specified. With regard to the update of the sample, it is vital to indicate if the sample is a fixed panel, rotating panel or entire sample renewed at regular intervals or other (for further details about sampling methods see chapter 4)

9.11 In *sample size*, the predefined total number of responses to obtain in each survey wave have to be stated.

9.12 *Sampling error*. Here, the desired precision of the estimates (for details see chapter 5) should be indicated. For tendency surveys, in most cases estimates are proportions (or percentages) of individuals expressing positive/negative opinions: the level of precision may be expressed in terms of percentage points at a fixed confidence level.

9.13 In the point *response rate,* the response rate expressed in the form of percentage has to be made explicit. For BTS, it can be useful define response rates in terms of weighted responses because for these surveys the value added/production is more interesting than the number of responding firms (for further details about definitions of response rates see chapter 7).

9.14 In *treatment of non-responses*, methods used to correct the bias due to the non-response - with respect to both item non-response (missing value) and unit non-response (missing records) - have to be pointed out (for further details about methods suitable for business and consumer tendency surveys see chapter 7).

9.15 The *sample coverage* means the extent or area covered by the sampling. It may be calculated in terms of units (expressions such as "50% coverage" means that one-half of the population under discussion have been examined in the sample), number of employees, amount of turnover etc. For BTS, the sample coverage in terms of employees or of the amount of turnover may be useful.

9.16 In the item *weighting scheme*, the weights used in the process of integration from the sample to universe figures have to be made explicit (for details on the choice of weights and process of estimation for business and consumer tendency surveys see chapter 5).

9.17 In *sectors and/or categories currently covered*, the list of economic sub-sectors (for BTS, according to an official statistical classification, e.g. divisions, groups of NACE rev.2) or categories (for CTS, in terms of income classes, age groups etc.) covered by the survey and the available sectorial breakdowns have to be stated.

9.18 In the item *periodicity*, the frequency of the survey has to be pointed out. It should be noted that some questions may have different frequency with respect to the survey (e.g. the survey has a monthly frequency but some questions are asked quarterly or biannual): the periodicity and the time during which these questions are asked should be indicated clearly in the metadata.

9.19 *Survey method*. The method used for data collection has to be described: CAPI, CATI, CASI etc.. For the most common methods used for business and consumer tendency surveys see chapter 6.

9.20 *Fieldwork period*: indicate here which days of the period the fieldwork is carried out (e.g. for monthly surveys: 1-20 of each month, etc.).

9.21 *Remarks methodology*. This item may be used to indicate remarks about methodology.

**(iii) Data and dissemination**

9.22 In data and dissemination, additional information about data, dissemination of results and release policy are requested.

9.23 The *unit of measure* of each indicator have to be reported here. For business and consumer tendency survey, most indicators are qualitative and the unit of measure is the percentage (proportion) of individuals.

9.24 *Data revision*. Periodic (e.g. for seasonal adjustment) or exceptional revisions (they may happen e.g. after new estimates due to a change in statistical classifications of economic activities or to a regular updating of weights used in data processing) may be specified in the item.

9.25 The *frequency of dissemination* should be indicated: monthly, quarterly etc.

9.26 *Dissemination format*: all the information concerning formats used for the data dissemination have to be described: publications, online database, micro data access, or other (e.g. website address etc.).

9.27 In the item *Release policy*, information about release calendar (specify when the calendar is updated) and users access to the calendar may be described.

9.28 *Metadata update*: the date of metadata last certified, last posted and last update should be stated.

**B. Publication procedures**

**(i) Variety of users**

9.29 The quality of statistical output is determined by several factors. Ultimately, the users of the output decide on the quality. The results of tendency surveys may attract a variety of users, including respondents to the survey; business executives; government policymakers; staff in banks and financial institutions; economists; researchers and analysts; press and other media. These (potential) users of tendency surveys do not all seek the same type of information. Some users analyse the statistics in detail while others only view the main results. This calls for a multiple approach to survey data publication.

**(ii) Quality criteria**

9.30 Following publication procedures with focus on quality criteria can improve the output and the satisfaction of all potential users. Timeliness, comparability, transparency, relevance, accessibility, interpretability and confidentiality are important quality indicators for the publication of business and consumer tendency surveys results. In general, quality is much enhanced by standardisation of publications in terms of time (timeliness, periodicity, simultaneous release to all interested parties), form (press release, analyst report, tables, database) and place (same location on a website).

**(iii) Release calendar**

9.31 Releasing publications and results according to a yearly calendar disseminated well in advance is an obvious way to increase the use of tendency surveys. It also helps to ensure periodicity and timeliness. It is important to release the survey results as soon as possible because timeliness is one of the main features of tendency surveys compared to other economic statistics. In the Joint Harmonised Programme of the EU, data collection for the monthly surveys takes place in the first two weeks of the month. The results are published on the second last working day of each month. The results of the quarterly survey are reported in the first month of the quarter (i.e. January, April, July and October). The release calendar is made public on the website in advance. Standardised publications are being released in the same way in every reporting period, which is convenient for users as well as statistical offices.

**(iv) Forms of dissemination**

9.32 Dissemination of business and consumer tendency surveys can take several forms, including publications like press releases and more detailed analyst reports; graphs and tables; downloadable database and extra information on request. Furthermore, metadata (see section A) should always be available as auxiliary information to increase the transparency and usefulness of the data.

9.33 The usage of tendency surveys and derived composite sentiment indicators is driven by increased demand for fast economic information and by digital publications that are more integrated and interactive. The publications combine timeliness with a sufficient detail of information. A shift in emphasis is occurring from reporting numbers to offering presentations and analytical tools. With different types of output, the heterogeneous interests of users should be met as closely as possible.

9.34 *A press release* should be short, at most one or two pages. Only the main results of the surveys are covered in press releases. The results of surveys can be summarised in composite

indicators. An overview table with key indicators and a graph showing the overall development of a few indicators are also useful. The press release includes brief methodological information about the number of respondents and coverage of the survey. A telephone number for press contacts should be given. In the Harmonised Programme, two separate press releases are sent out to news agencies for the Economic Sentiment Indicator (ESI) and the Business Climate Indicator (BCI).

9.35 More information can be published in an *analyst report*, which accompanies the press release(s). The target audience includes every user seeking more detailed results. Besides describing the main results, the report presents detailed data in tables for all variables by sectors; regions; etc. These tables should show quarterly or annual data spanning for the recent past and some other statistical indicators as long-term average and min/max values. In the Harmonised Programme, analyst reports are being published simultaneously with the press releases for the main indicators, in the form of annex tables attached to the ESI press release. These tables cover composite indicators and balance results per question and per country for the past 12 months, along with information on historic minimum, maximum and average values.

9.36 *Special reports* can be sent on demand to respondents to the survey or other parties if allowed by the statistical law. Special reports with detailed results on the specific branch individual respondents are working in are an important tool to incentivise managers to respond to the survey Interested respondents can benchmark their own results with their sector's. Apart from the dissemination of the pure survey data, analytical reports using the survey data as input for more general (socio-)economic analysis or forecasting can contribute to the wider dissemination of the survey results and can raise awareness of and demand for the data (with possible positive effects also on response rates). For instance, DG ECFIN publishes a quarterly analytical note employing survey data ('European Business cycle Indicators') on its website.

9.37 *Graphs, tables, database.* Besides the headlines presented in press releases, there seems to be great variation in the information users are interested in. Some are interested in sectorial or regional splits, others are even interested in specific questions like on employment, or mainly in forward looking questions. It is therefore useful to publish all the results of the surveys. In this spirit, detailed results can be published simultaneously with the press on website in the form of tables containing long time series on all composite indicators, underlying balances, including breakdowns for industrial sub-sectors or consumer categories. Composite indicators are usually the best predictors tendency surveys can produce with respect to the actual economic development. Individual indicators and answers to separate questions are more volatile and must be interpreted

with more care. Even if individual series are seasonally adjusted, they can be difficult to interpret. Clear descriptions and extra explanation of the data should therefore accompany the tables and database.

9.38 The main channel to publish results of the business and consumer tendency surveys is the *website.* Results can be published quickly online and are available to everyone at the same time. On the website, a statistics database with adjustable tables, graphs and extensive methodological information can also be made available. The development and maintenance of a database for online data sharing requires specialised software. Different types of users can access information to their needs. The design of the website should meet many requirements. Routing on the website can become complex when large amounts of data and information are offered. Interpretation of survey results can also be difficult. Interpretability is of concern in the case of tendency surveys results. Balances can be difficult to interpret and compare across sectors and regions, even for regular users of the data. Therefore, a link to a comprehensive metadata file explaining the underlying concepts and definitions is vital. Websites offer the possibility of more interactive and dynamic (animated) applications to present the results of the surveys.

**(v) Confidentiality**

9.39 Individual data gathered with business and consumer tendency surveys are largely confidential. Publication of the data is therefore bound to limits. In CTS, official publications do not contain confidential data, in BTS the data producers must be on guard not to disclose too much information. Especially when both regional and sectorial data are published, there is a risk of exposing information of individual businesses.

Both primary and secondary confidentiality should be guarded to avoid unwanted revelations. Primary confidential is data of a unit which can be directly identified. Secondary confidential is information on a unit that can be identified when combined with other data. If micro data is made available for (academic) research purposes, it should always be made anonymous.

**(vi) Revisions**

9.40 There is a trade-off between timeliness on the one hand and accuracy and level of detail of research results on the other. As timeliness is a key quality of business and consumer tendency surveys, sometimes revisions are necessary. Extra information from incoming data usually does not

lead to revisions of tendency surveys data.[31] Revisions for other reasons can however take place, planned or unplanned. For instance, seasonal adjustment can cause revisions, depending on the applied method (see Chapter 8). Regular updating of country weights in the computation of regional aggregates also inevitably leads to some revisions. A change in the Statistical classifications of economic activities) classifications is another circumstance which can make revisions necessary. Like in all statistics, mistakes in published data may also require revisions. In the EU Programme, no general harmonised regulation exists for revisions of BCS data, but obviously they must be described and communicated well in all cases.

[31] An exception is, for example, the French survey programme ran by the INSEE, where late replies (received after the month's cut-off date) are included in the following month's results, thereby leading to (slight) revisions.

## References

European Commission, Directorate-General for Economic and Financial Affairs (2014), *"The Joint Harmonised EU Programme of Business and Consumer Surveys"*, user guide March, available at http://ec.europa.eu/economy_finance/db_indicators/surveys/method_guides/index_en.htm.

EUROSTAT, *Short term business statistics* Available on: http://epp.eurostat.ec.europa.eu/cache/ITY_SDDS/en/sts_esms.htm.

OECD (2003). *Business Tendency Surveys: A handbook.* OECD Publications, Paris, ISBN 92-64-19894-6. (http://www.oecd.org/std/leading-indicators/31837055.pdf.

# CHAPTER 10

## Use of tendency survey results

### A. Users of survey data

10.1 At their origin, the basic purpose of ETS was to collect information on business conditions for the benefit of respondents; hence, surveys were carried out mainly by trade associations. Nowadays, ETS are considered as an important tool for economic surveillance, short-term forecasting and economic research; as such, a large number of private research institutes, official statistical agencies, ministries, central banks and national and international government bodies are involved in the collection and the analysis of these data. ETS results are also broadly referred to in generic and specialized media and are closely monitored by financial analysts.

10.2 More specifically, existing literature identifies four categories of potential survey users (OECD, 2003):

a) international bodies responsible for economic surveillance - such as the European Central Bank (ECB), OECD or EC
b) national policymakers - such as National Central Banks, responsible for monetary policy, finance ministries/treasury departments - responsible for fiscal policy, other ministries/government departments dealing with employment and industrial policies, administrative users such as national statistical offices
c) data providers, such as economic research institutes, businesses and their representatives such as industry/trade associations and chambers of commerce, analysts in general, financial institutions
d) the press and other media.

Moreover, potential users may also be distinguished in two broad categories according to their different needs: those who intend to analyze the statistics in detail and those who simply want to know the main results. The first category – "analysts"– includes economists and researchers operating in the academia, research institutions and governmental bodies. The second category may be labeled "executives" and consists of senior business executives, politicians, senior civil servants responsible for government policy, senior personal in banks and financial institutions and the press and other media. "Executives" probably make up the majority of BTS data users.

10.3 At the international level the EC may be considered both as an analyst and an executive user of ETS data. The EC has a long history of promoting the harmonization of EU member states surveys to such an extent that results are comparable and aggregated at the EU level. Within the EC, the main 'executive' uses of ETS are the continuous monitoring of the current economic situation in both individual Member states and the euro area and EU aggregates. ETS data also regularly feed into nowcasting and short-term forecasting exercises, including the European Economic Forecasts – published in Spring, Autumn and Winter of each year. These macro-economic forecasts contain inter alia projections of GDP growth and consumer price inflation for the euro area and the EU; they are also used as input into the EC economic surveillance tasks in the framework of the Stability and Growth Pact. Other important users among the international governmental organizations are the OECD and the ECB. The OECD derives for almost every member states an indicator based on survey data with good leading properties; it also picks headline indicators (for industry and consumers) from the business and consumer tendency survey database (and from other sources for non-EU countries) and transforms them to create the Composite Leading Indicators. OECD goes into question level detail, for example order books, employment but does not use survey data for the sub-sectors. It also regularly publishes survey results in conjunction with "hard data" in the Main Economic Indicators database. Also the ECB is a heavy user of survey data, being particularly interested in questions relating to price developments. Similarly, questions on developments in output, demand and employment as well as capacity utilization or more structural information (such as firms' investment plans ) are considered quite relevant for back-casting/nowcasting/forecasting GDP (see below), for business cycle analysis and macroeconomic surveillance.

10.4 ETS data are also used by national policy makers and administrative users to monitor the current economic situation and to get an advance indication of the direction in which the economy is heading. Among them, the Conference Board and the University of Michigan publish and analyse indicators referred to the US economy, while the Bank of Japan conducts the Tankan survey for the Japanese productive sector.

10.5 Executive users include research institutes (both public and private), and private sector organizations such as financial institutions, industry associations and the media; among them, the ESI and the Consumer Confidence Indicator (CCI) are the most popular products (see below in this chapter for a description of these indicators). Private sector users may be above all interested in questions including financial situation, savings, competitive position, production levels and

expectations, orders and employment; private sector users tend to use ETS for monitoring the current economic situation and for getting a general overview of business/economic conditions. More analytical users such as data producers, large financial institutions and research institutes put survey products to quantitative uses in forecasting and macroeconomic modeling; in some cases, particularly among financial institutions, ETS data are also used to inform business/investment decisions. Finally, last, and not least, the biennial CIRET conferences may be considered as a good expression of the broad worldwide dissemination of survey data.

**B. The use of survey data**

10.6 The surveys are implicitly based on the assumption that insiders are in a privileged position to assess developments in a given market; therefore the same insiders, if taken on an aggregate basis, can roughly anticipate business cycle trends. The surveys came into vogue as a purely empirical instrument to track the business cycle with the big post 1929 recession and were further developed after the Second World War as an early warning tool to manage anti-cyclical policies. All in all, there is no clear economic theory explaining why the surveys should work and the subject has been still nowadays the matter of a fierce debate among the economic scholars. Nevertheless, there is a quite copious amount of well documented studies that recognized survey data as a crucial instrument for gathering information in today's ever-changing environment (see among the others, Koopmans, 1947; Zarnowitz, 1992).

10.7 From a general standpoint, over the years survey data have been mainly used:

a) for constructing business cycle composite indicators with various methodologies, ranging from the classical heuristic NBER approach to factor models (FM) and markovian models
b) as visualizing tools of the business cycle situation
c) as input variables in short-term nowcasting and forecasting models
d) in macroeconomic analysis
e) as a source of information for micro-econometric analysis of agents' behavior.

10.8 Nevertheless, this classification provides a first and useful introduction to existing literature in this field, not to be intended as an all-comprehensive account of the literature produced over the years.

**(i) Business cycle composite indicators**

10.9 Classifying economic variables into leading, coincident and lagging indicators of economic activity is a long-lived tradition in economic research (Burns and Mitchell, 1946). According to the now classical NBER (National Bureau of Economic Research) definition, business cycles are recurrent sequences of alternating phases of expansion and contraction in economic activity. However, the term "business cycle" has some ambiguity, since it can refer to conceptually different economic fluctuations; it can be referred to the "classical cycle" when focusing on fluctuations in the level of the economic activity; "growth cycle", when concentrating instead on fluctuations in the economic activity around the long-run potential level and to "growth rate cycle", when interesting in fluctuations of the growth rate of economic activity. ETS data are typically considered in a growth cycle or growth rate cycle framework because of their trend free nature. The advantage of composite indicators over the individual component series is that they achieve a better trade-off between responsiveness and stability. Composite indicators can be constructed to have fewer false alarms and fewer missed turning points than its individual components; moreover they tend to have more stable lead-times. Finally, the composites have the capacity to react to various sources of economic fluctuations and at the same time can be resilient to perturbations affecting only one of the components.

10.10 Generally speaking, three procedures are used to build composite indicators: the traditional heuristic NBER approach, the factor model approach and the markovian methodology. The set of survey variables used to compute composite indicators and the way to aggregate them depend on the goal assigned to the indicators.

10.11 The heuristic approach has its roots in the standard NBER methodology and it is based on a high degree of experience of the researcher in charge of the analysis. In this approach, still vastly used in Europe and in the US, variables are selected "by hand" relying on descriptive statistics and on the detection of turning points. Once relevant variables have been selected, a simple way to aggregate them is to compute the arithmetic mean of net balances, as it is the case for the computation of the confidence indicators; a weighting scheme may also be chosen to aggregate between different sectors and, if needed, between different countries.

10.12 Within this approach, survey producers in EU countries have designed confidence indicators maximizing the coincident correlation with the reference series (for instance the year-on-year growth rate of industrial production in the case of the industry survey); hence, the harmonized

confidence indicators computed in EU countries can be considered as composite coincident indicators for each sector.

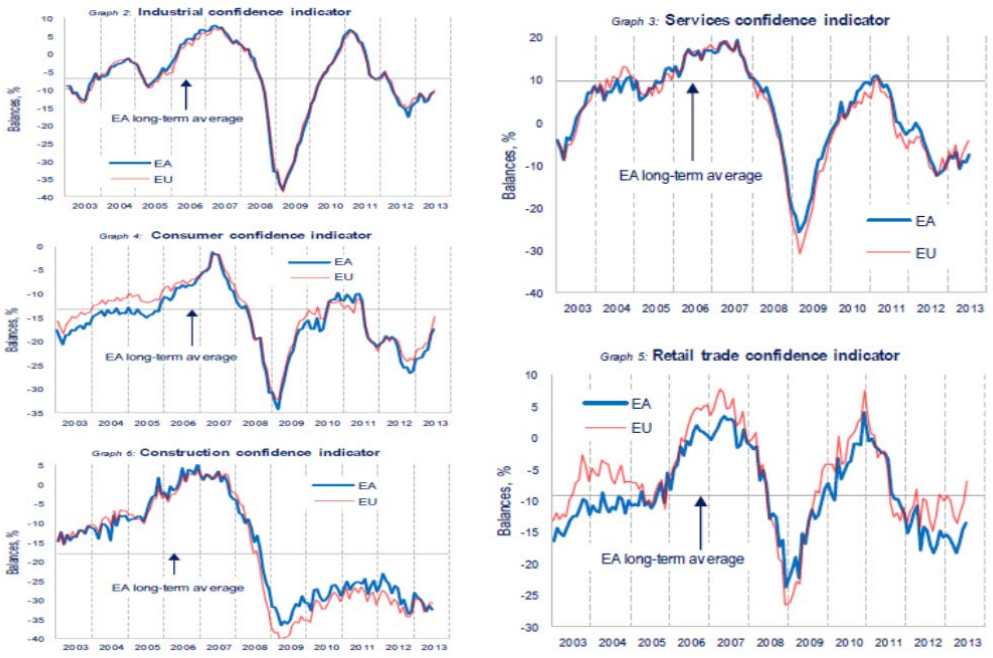
10.13 More specifically, five separate confidence indicators for each surveyed sector in individual EU countries are elaborated, considering the following variables (EC, 2014):

- The industrial confidence indicator (ICI) is an average of the balances of questions in the industry survey relating to production expectations, total order books and stocks of finished goods (with the inverted sign);
- The construction confidence indicator is an average of the balances of questions in the construction survey relating to total order books and employment future tendency;
- The retail trade confidence indicator (RCI) is an average of the balances to the questions in the retail trade survey relating to present business situation, future business situation and stocks (with the inverted sign);
- The services confidence indicator (SCI) is an average of the balances to the questions in the survey relating to present business situation, past demand and future demand;
- The consumer confidence indicator (CCI) is the arithmetic average of the balances to the questions on the financial situation of households, the general economic situation, the unemployment expectations (with the inverted sign) and savings, over the next 12 months in each case.

10.14 The EC calculates then the EU and euro-area aggregates on the basis of the national results and seasonally adjusts the balance series (see figure F10.1). Notwithstanding the simplicity of the method used to construct confidence indicators, it performs rather well in general (see on this issue Gayer and Genet, 2006); a possible explanation for such a good performance is that this type of aggregation is similar to pooling in the forecasting literature, which is known to work quite well in general (Clemen, 1989; Stock and Watson, 1999). Moreover, the rationale for the construction of such indicators is justify by the circumstance that a sectorial analysis may be very interesting for a better evaluation of business cycle conditions, for example for monitoring sectors that typically anticipate the global cycle such as building or that represent important engines of growth as services or consumption from a demand perspective.

**Figure F10.1**

**Confidence indicators**

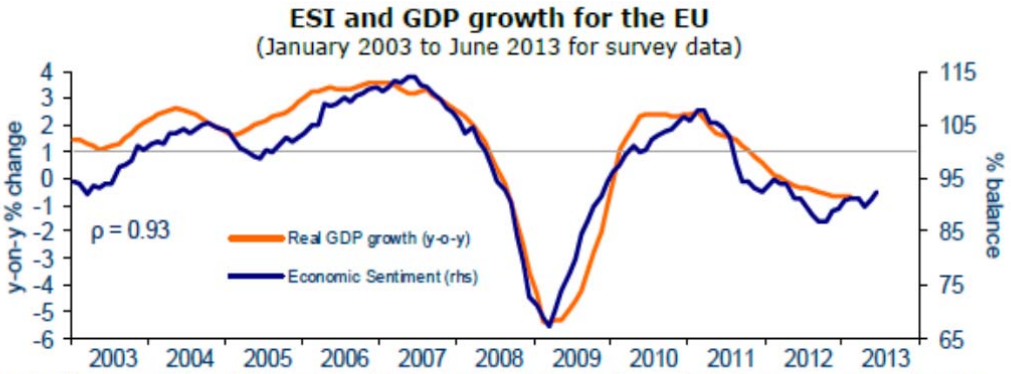



**Source: European Commission, Business and Consumer surveys**

10.15 Accordingly, alongside these indicators, the EC also provides the so-called ESI as a weighted average of the balances of replies to some selected questions addressed to firms and consumers in the five sectors covered by the EU Business and Consumer Surveys Programme (consumers, industry, construction, services, and retail). The EC monthly calculates EU and euro-area aggregates on the basis of the national results and seasonally adjusts the balance series. The indicator is scaled to have a long-term mean of 100 and a standard deviation of 10; hence, values greater than 100 indicate above-average economic sentiment and vice versa. This indicator – drawing on confidence indicators in different sectors of the economy – can be considered very effective as a valuable tool to now-cast GDP before it is released by national accounts (see figure F10.2).

**Figure F10.2**

**ESI**




*Note 1: The horizontal line (rhs) marks the long-term average (=100) of the sentiment indicator.*
*Note 2: Both ESI and y-o-y GDP growth are plotted at monthly frequency. Monthly GDP data are obtained by linear interpolation of quarterly data.*

**Source: European Business Cycle indicators**

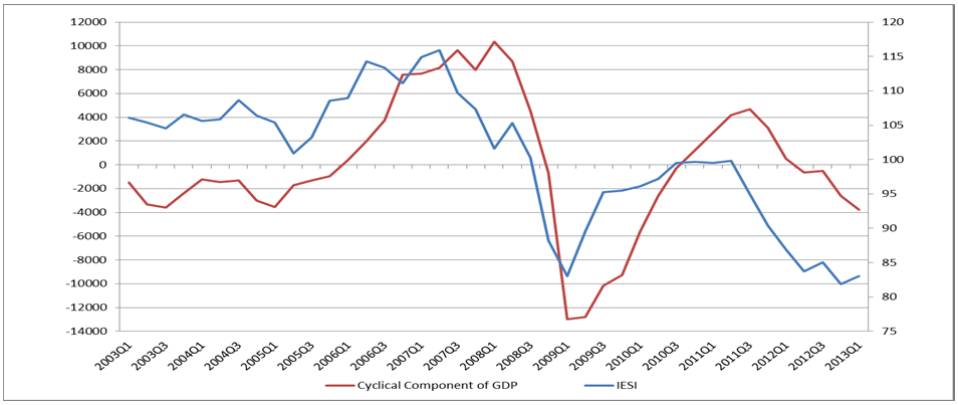
10.16 Furthermore, alongside the business cycle indicators, a special and prominent importance is devoted to leading indicators, mainly because to their potential contribution to business cycle forecasting. According to De Leeuw (1991), variables to be used for constructing leading indicators should be: timely (there are some lags between production decisions and actual production); easy to adapt (some variables adjust more slowly than others to the same economic stimulus); related to market expectations (some series are influenced by expectations of future changes in economic activity); price movers (some variables, such as monetary and fiscal policy instruments). However, whereas a set of theoretical rationales can be put forward to constitute a leading behavior of some economic indicators, empirical studies about the use of such indicators in forecasting are quite ambiguous and have sometimes contradictory conclusions (Emerson and Hendry, 1996).

10.17 Anyway the relevance of such indexes encourages almost all the countries to provide and disseminate leading indicators. For example, a prominent leading indicator for the business cycle in Germany is the IFO Business Climate for industry and trade (aggregating results for the manufacturing, construction, wholesaling and retailing sectors); it is computed as a geometric mean of the balances referring to the current business situation and the business outlook in the next six months. Another example is the Italian ISTAT Economic Sentiment Indicator (IESI), which showed itself to be able to anticipate the fluctuation of aggregate economic activity. The IESI is based on the set of balance series underlying the confidence indicators of industry, construction, services and

retail trade, normalised and weighted using value added shares. Apart from the exclusion of the results from the CTS, it actually follows quite faithfully the EC methodology used for the ESI elaboration. The Italian GDP is chosen as the reference series to evaluate the properties of the indicator; in particular, the relationship between the ISTAT Economic Sentiment Indicator and the Italian GDP is evaluated looking at the turning point coherence, the correlation and the directional coherence (see figure F10.3).

**Figure F10.3**

**IESI and cyclical component (Hodrick-Prescott) of GDP**

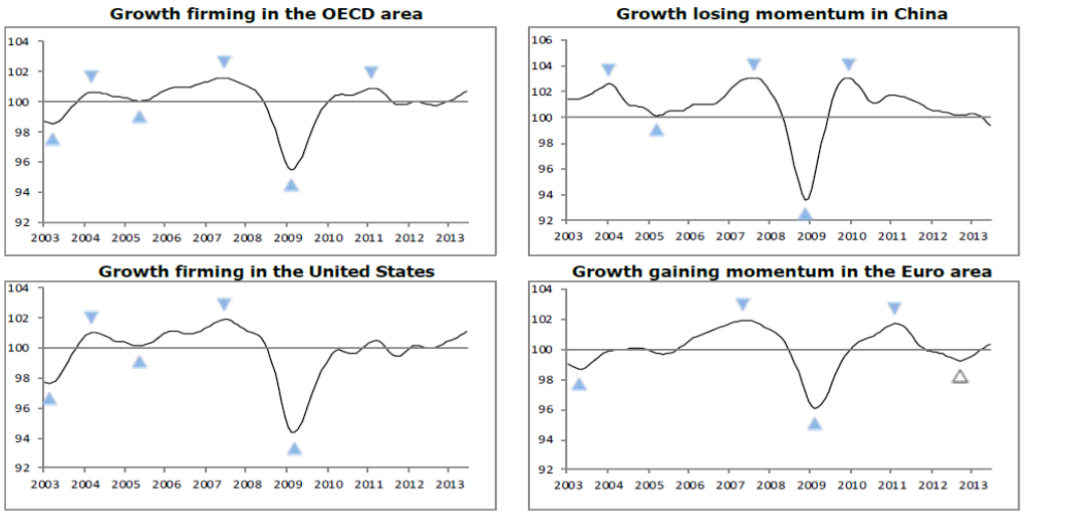



**Source: ISTAT**

10.18 Finally, BTS data can be also combined with quantitative statistics to obtain more structured cyclical indicators. For example, the OECD system of CLIs is designed to provide early signals of turning points in business cycles, considering fluctuations of economic activity around its long term potential level. This approach, focusing on turning points (peaks and troughs), provides qualitative rather than quantitative information on short-term economic movements. CLI is constructed by aggregating together component series selected according to multiple criteria, such as: economic significance, cyclical correspondence and data quality. The phases and patterns in CLIs are likely to be followed by the business cycle (see figure F10.4).

**Figure F10.4**

**OECD Composite Leading Indicators**




The above graphs show country specific composite leading indicators (CLIs). Turning points of CLIs tend to precede turning points in economic activity relative to trend by approximately six months. The horizontal line at 100 represents the trend of economic activity. Shaded triangles mark confirmed turning-points of the CLI. Blank triangles mark provisional turning-points that may be reversed.

**Source: OECD Composite Leading Indicators News Release**

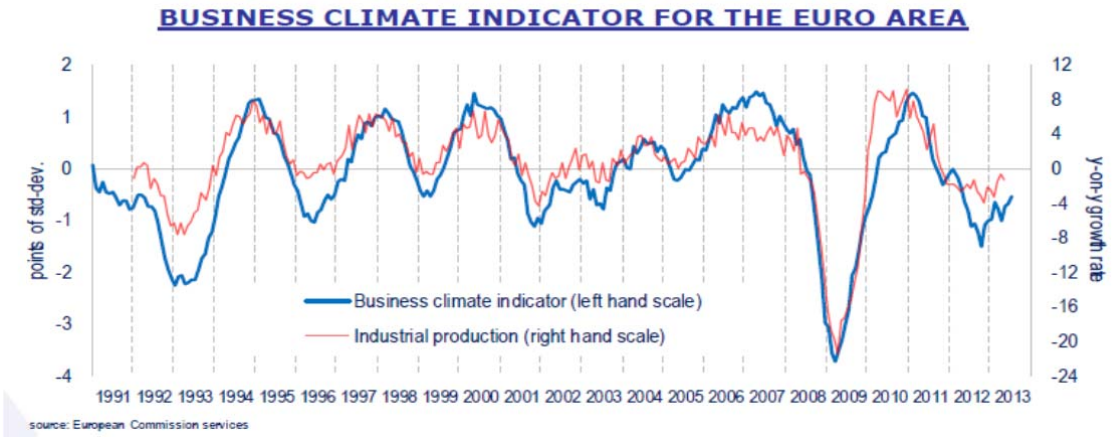
10.19 In addition to the heuristic "model-free" composite indexes, recent advances in the cyclical analysis – partly stimulated by a set of theoretical developments - provide alternative empirical methodologies to construct a new generation of composite indicators based on a model-based framework: FM and Markov Switching Models (MSM). The general idea is to combine several indicators in order to derive a cyclical signal, smoothing the original series and cleaning them from idiosyncratic noise. In both cases there is a single unobservable force underlying the current state of the economy, but in the former method this is a continuous variable, while in the latter one it is a discrete variable that evolves according to a Markov chain. In this sense, the factor model approach can be seen as a formalization of Burns and Mitchell's (1946) notion that business cycles represent co-movements in a set of economic time series, while MSM are well suited to capture the different behavior of the variables during expansions and recessions.

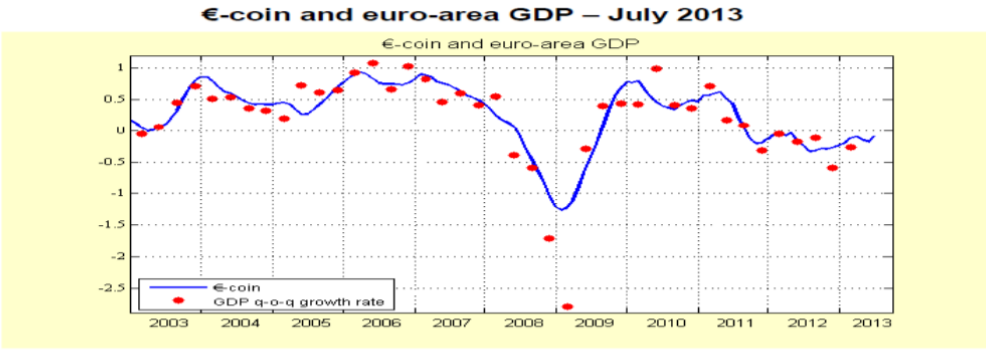
10.20 Widely used empirical methodologies are developed along the factor model-based methodology; the two most common techniques are suggested by Stock and Watson (2002 a,b) in the time domain and by Forni et al. (2005) in the frequency domain. Basically, the first methodology suggests estimating the common factors as the static principal components of the

variables, while the other one proposes the use of dynamic principal components to better capture the dynamics of the underlying time series. As an example of composite indicators using factor methodologies, the EC calculates the so called Business Climate Indicator - elaborated exclusively for the industry sector of the Euro Area - while Bank of Italy and CEPR monthly publish the Eurocoin indicator which includes - among the others - survey data in order to assess the current stance of the European business cycle. Also national institutes responsible for the realization of ETS often make use of FM in order to convey early signals on business cycle developments, basing their analyses exclusively on survey data; for example the INSEE composite business climate indicator is compiled on the basis of 26 balances of opinions from the manufacturing, services, building, retail and wholesale trade's surveys indicators (see on this issue the examples displayed in figure F10.5).

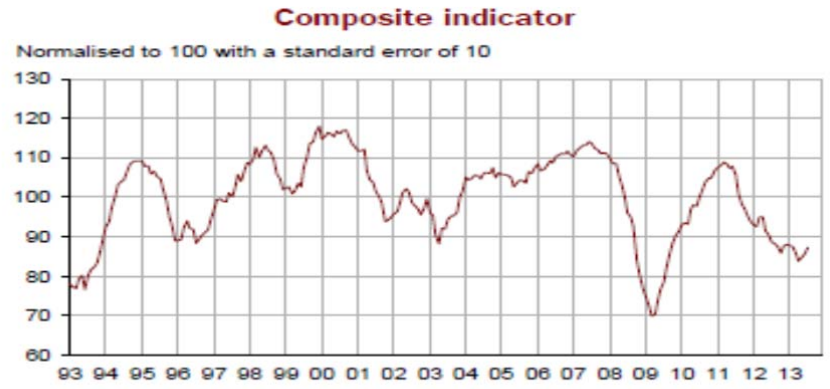
**Figure F10.5**

**Examples of factor-based composite coincident indicators**




*Sources:* Bank of Italy and Eurostat.




*Sources : Insee, business surveys*

10.21 Since FM are particularly suited to extract summary information from large datasets, they represent an alternative natural statistical tool to derive model-based composite coincident indicator using survey data. Gayer and Genet (2006) compare the model-free procedure followed by the EC with several factor-based composite coincident indicators, finding that the more sophisticated methods yield no gains for the industry sector and limited gains for the other sectors. However, alongside more specific studies about dating cycle analysis, Hamilton's (1990) MSM provides a more convenient statistical framework. Basically, the composite indicator in this context represents the status of the business cycle (expansion/recession mapped into a binary variable), which determines the behavior of all the coincident indicators that can change substantially over different phases of the cycle. Diebold and Rudebusch (1996) suggest that these two approaches can be combined by allowing the underlying factor in the Stock and Watson model to evolve according to a MSM.

**(ii) Visualizing tools**

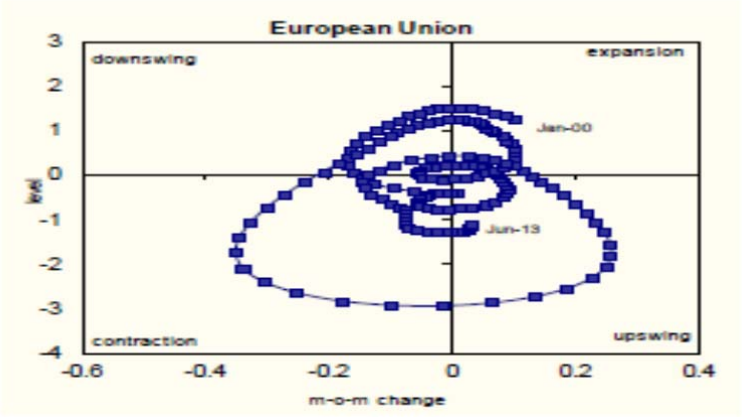
10.22 Recently, business cycle indicators are also used to convey an immediate, self-explicative visual representation of the business cycle situation, through the construction of the so-called visualizing tools. The EC produces the "Economic Climate Tracer" (Gayer, 2010) aimed at visualizing the state of the economy using sectorial confidence indicators; the indicator is directly derived from the Joint Harmonised EU Programme of Business and Consumer Surveys and is included in the Commission's quarterly publication 'European Business Cycle Indicators'[32]. It is based on smoothed confidence indicators and plots their levels against their month-on-month changes, resulting in circular, counter-clockwise movements through the four quadrants of the

[32] http://ec.europa.eu/economy_finance/publications/cycle_indicators/

graph. The latter directly corresponds to a turning-point oriented definition of the business cycle phases (figure F10.6).

**Figure F10.6**

**Economic Tracer**

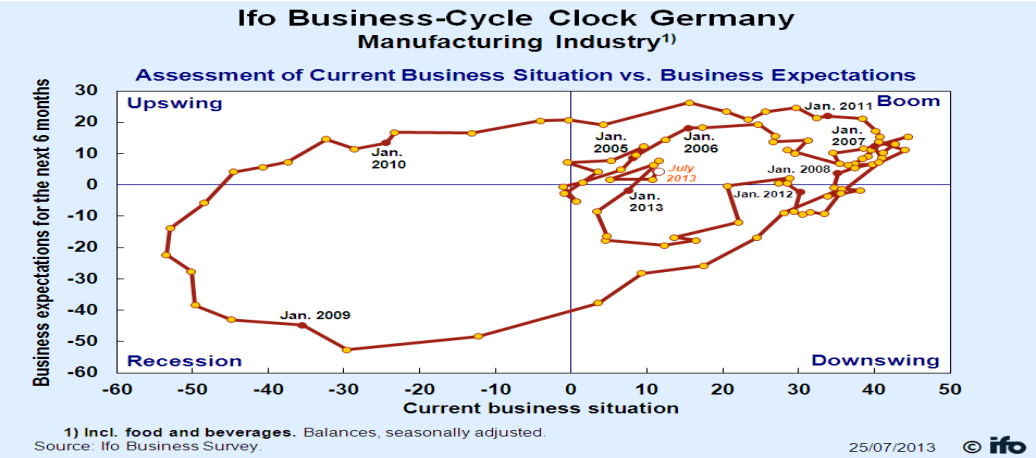



**Source: European Business Cycle indicators**

10.23 Various other graphical monitoring systems that display the cyclical component of selected indicators and their modification (vis-à-vis the previous period) in a 4 quadrant scheme have been recently developed, including the OECD Business Cycle Clock, the Eurostat European Business Cycle Clock, the Business Cycle Tracer of the Central Statistical Office of the Netherlands and the IFO Institute's Business Cycle Clock. The latter (figure F10.7) visualizes the interaction between managers' business assessment and expectations (see on this issue Ruth, Schouten and Wekker, 2005; Eurostat, 2010; Abberger and Nierhaus, 2010).

**Figure F10.7:**

**IFO Business Cycle Clock**

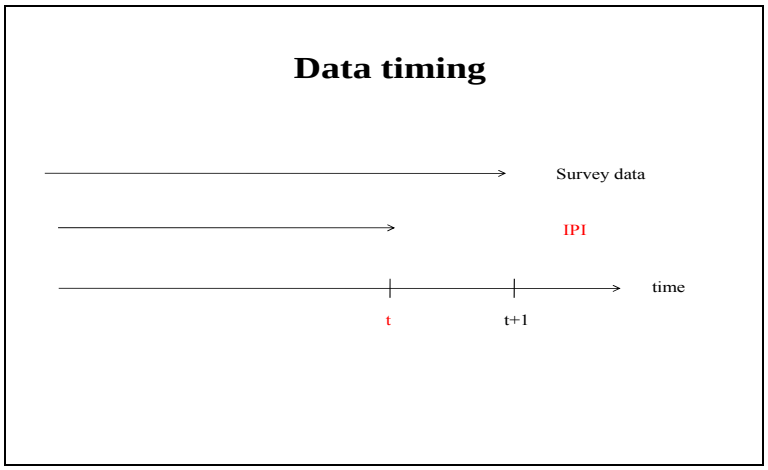



### (iii) Forecasting using survey data

10.24 From a general point of view, the rationale for selecting indicators to be included in forecasting models is essentially based on their statistical characteristics: in this sense, timeliness and absence of revision, immediate reliability and a significant correlation with the economic variable to be forecasted can be considered ideal requisites for forecasting purpose. Several studies find that forecast accuracy improves with the inclusion of ETS data (see for instance Diron, 2008; Giannone et al. 2009). Furthermore, survey questions are phrased in such a way to allow as much forward-looking insight as possible, reflecting agents' expectations which are likely to bear a relation to future developments of macroeconomic variables. As a consequence, ETS data are widely used as potential predictors of some relevant economic variables such as GDP, consumers' expenditures or industrial production. In this context, the different timing between survey data and the variable to be forecasted has a prominent role. As an example, considering the Italian index of industrial production (IPI) published monthly by the National Institute of Statistics (ISTAT), survey data are released about 45 days before the industrial production index; so when the latter is available for month t, the former is available at least up to month t+1 (see figure F10.8 below). Therefore, even if the survey indicator is strictly coincident, its use allows a one-step ahead forecast of the industrial production index.

**Figure F10.8**

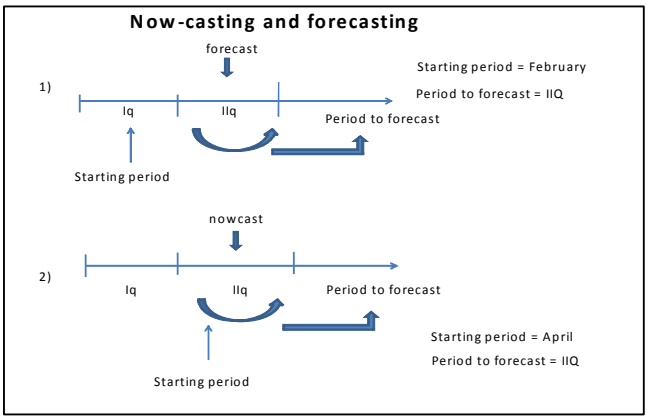
**Data timing**




10.25 A number of econometric tools have recently been developed to predict economic behavior in the current and immediately following quarters; these kind of exercises are commonly referred to as nowcasting and short-term forecasting respectively, (see figure F10.9 for a graphical description). However, there is no general consensus in the empirical literature about the best method to use; all the techniques and the empirical applications could be indeed attractive and promising, depending on the starting hypotheses to be tested, the targets to be reached, the variables to be forecasted, the forecast horizon to be considered (short-term or medium-term).

**Figure F10.9**

**Nowcasting and forecasting**




10.26 For example, as far as indicators' selection is concerned, the activity of extracting reliable signals from high frequency indicators can be carried out according to two main directions: the FM

and the Bridge Models (BM). As said before, FM have the advantage of summarizing all the available information into the extraction of a few common factors from a large set of indicators (Angelini et. al, 2011), while BM are a relatively simple and popular method which links the forecast targets to suitable indicators, selected a priori on the basis of the researcher's experience and statistical inference (see for instance, Golinelli and Parigi, 2007; Baffigi et al. 2004; Bruno and Lupi, 2004). Since the BM require that the whole set of regressors should be known over the projection period, these kind of models may be more properly conceived as a tool providing an early estimate of the current situation, a "now-cast" rather than a true forecast. However, in spite of their large usage, both FM and BM approaches are subject to criticism. FM may be biased by unbalanced sources of information (Boivin and Ng, 2006), while BM may appear excessively ad hoc because of the excessive discretion underlying the choice of list of the pre-selected indicators and consequently, the specification of the forecasting equation in the BM may be often quite difficult to replicate.

10.27 Finally, absence of revision makes ETS also useful in a real-time framework, i.e. considering the information set available to economic agents (typically, monetary of fiscal authorities) at the moment they are taking their decisions (for a broader definition of real time data, see for instance Crushore and Stark, 2001). More specifically, ETS may be very helpful when the variable to be predicted is prone to large revisions over time; in this context, survey data have been proved very useful to improve output gap estimations (Graff and Sturm, 2012), but not particularly effective to help predicting consumers spending (Crushore, 2005).

**(iv) Consumer confidence, macroeconomic analysis and business cycle**

10.28 The role of confidence indicators in economic analysis has been widely debated over the last years. The various approaches used in the literature range from the conclusion that confidence indicators have a crucial role in predicting macroeconomic variables and in signalling early business cycle turning points to strategies using econometrics techniques that consider that they have no role even in forecasting. In the following, a brief overview of the existing literature is provided; this is to be intended as an useful introduction to the debate and a possible starting point for further research, without any ambition of exhaustiveness.

10.29 Broadly speaking, the link between confidence and economic decisions has been usually covered focusing on two main aspects. From the theoretical point of view, the literature has mostly concentrated on the interpretation of confidence and its role in the consumption theory. Essentially, the Permanent Income Hypothesis (PIH) does not allow for a predictive power of the confidence index over consumption, beyond its capacity to signal changes in permanent income, while the studies in psychological analysis of consumer behaviour underline the importance of household's attitudes as independent causes of consumer spending (Katona, 1975). In this sense, consumer confidence is supposed to mirror aspects of consumers' sentiment which can have an impact on their propensity to consume (Roos, 2008).

10.30 From an empirical point of view, the literature has instead investigated the extent to which confidence indicators may contain any information over and above economic fundamentals. In essence, the hypothesis is whether confidence can be explained by current and past values of macroeconomic variables such as income, unemployment, inflation or interest rates (Lovell, 1975, Golinelli and Parigi 2004) or whether confidence measures may have any statistical significance in predicting economic outcomes even after appropriate macroeconomic variables are taken into account. Some studies suggest the importance of some political events as wars or elections, media coverage or some international and/or socio-political factors as possible determinants of consumer perceptions and sentiment (Vuchelen, 1995; De Boef and Kellsted, 2004; Malgarini and Margani, 2007; Ramalho et al., 2013).

10.31 Empirical results on effective forecasting ability of consumer confidence are however mixed and still inconclusive (see among the others, Matsusaka and Sbordone, 1995; Ludvigson, 2004; Malgarini and Margani, 2007). Notably, there is no unified approach to fully identifying and isolating the effect of consumer confidence on consumption. Indeed confidence is closely related to the economic variables that may affect household spending; consequently the seemingly conflicting results among the empirical studies may be linked to the different sets of economic indicators used in the various forecasting models (European Central Bank, 2013).

10.32 At the same time, another strand of the literature attributes a possible weakness in the disagreeing empirical investigations to the use of a linear functional form to model the relationship from consumer attitudes to confidence. The use of linear models, in fact, can be partially inconsistent with the view that only abrupt shifts in consumer confidence are relevant to signal changes in consumer expenditure. Also, there is some evidence that the relation between consumption and confidence may be unstable over time (Al-Eyd et al. 2008). By this interpretation,

some works have recently proposed the use of non-linear forecasting models to express the relation between confidence and household spending (see on this issue, Desroches and Gosselin, 2004). Bruno (2014) suggests a non-parametric non-linear approach allowing to avoid too restrictive assumptions about the functional form in order to explore the usefulness of confidence in forecasting consumption. Focusing on Italian data, asymmetries over the business cycle are found to characterize such a relationship, as the empirical results show that the expected consumption growth is always very negative as long as the confidence falls below certain threshold, while the converse in not true. Accordingly, using a different methodology, Dees and Soares Brinca (2013) find evidence about these asymmetric threshold effects for the euro area, supporting the idea of a non-linear relationship between confidence and economic fluctuations.

10.33 A major consensus about confidence indicators is instead achieved when the related role of consumer sentiment in business cycle fluctuations is considered. Indeed, by this interpretation, some authors stress the relevant importance of confidence in signaling early turning points (Howrey, 2001) and in predicting periods of strong fluctuations in the economy, such as recessions and recoveries or during periods of occasional events (such as the first Gulf war or September 11$^{th}$ attack, see on this issue Throop, 1992 and Garner, 2002, respectively). Acemoglu and Scott (1994) suggest that the circumstance that the UK consumer confidence predicts consumption even after controlling on income may be due to shifts of the consumption function over the business cycle.

10.34 One theoretical explanation can be found along the dynamic general equilibrium models that give rise to multiple equilibria, in which expectations about the future level of output can become self-fulfilling (Farmer, 1999). According to Harrison and Weder (2006), for example, agents' self-fulfilling expectations are one of the primary impulse behind fluctuations. In their paper, they find that these kind of shocks can well explain the entire Depression era, from the 1929-32 decline to the subsequent slow recovery and the recession that occurred in 1937-38. Empirically, the use of probit models has been widely used in modeling business cycle phases, in particular in predicting the likelihood of recessions and expansions some quarters ahead of the current period. Evidence shows that business and consumer tendency surveys are pro-cyclical and usually play a significant role in predicting recessions (Taylor and McNabb, 2007).

10.35 Some other authors also underline that the business cycle fluctuations can arise even without any significant changes in economic fundamentals and recessions can develop merely from a wave of pessimism following a negative confidence shock associated with an increase of uncertainty (Bloom, 2009; Bachman et al. 2013). This occurs because higher uncertainty causes households to

increase their precautionary savings (reducing equivalently their consumption expenditures) and firms to suspend their investments and hiring, affecting in turn consumer sentiment and spending. In the medium term, uncertainty shocks lead output, employment and productivity to overreact, generating sharp recessions and recoveries. In other words, changes in expectations driven by changes in sentiment may drive economic developments.

10.36 Recently, confidence indicators have also been used in the literature concerning international synchronization of business cycles, both at the global level (Kose et al., 2012) and the regional one (Kim and Kim, 2013). In Europe, the analysis has especially concentrated on the degree of convergence of new entrants with respect to member countries (Artis, Marcellino, Proietti, 2005) and on the effects of the monetary union on convergence (Aguiar-Conraria and Soares 2011). However, there is no widespread consensus on the degree of convergence reached at the global and local level; one of the main reason behind the difficulties experienced in reaching a consensus is that the analysis is usually based on trended data (typically, GDP, industrial production, trade volumes) that should be de-trended before the analysis. Results are hence particularly sensitive to the methods used to extract the cyclical component of trended data. In this sense, ETS data may be considered as ideal candidates to be considered in this type of analysis, since -as already pointed out - they contain no trend and hence do not need any pre-filtering. Using the EC Confidence indicators, Gayer and Weiss (2006) find some evidence of a recent decrease in Euro-area business cycle synchronization starting around 2002; similar results are also found in Gayer (2007). Recently, analyzing the EC confidence indicator and the US consumer confidence, – as measured by the University of Michigan - Dees and Soares Brinca (2013) show that there is not only a clear relationship between the US consumer confidence and the corresponding euro area index, but also some lead in the US sentiment when compared with the euro area, underlying the importance of the transmission of shocks from the United States to the rest of the world.

10.37 Finally, substantial discrepancies in activity at sectorial level justify the interest - along the business cycle analysis - in a sectorial disaggregation. In this context, ETS data may be also used to analyse changes occurring at the industry level (see Goldman Sachs, 2013); indeed, ETS may provide very useful information in this field, thanks to its timeliness (w.r.t "hard" data like those on industrial production) and availability of sectorial information (larger w.r.t other short term indicators like the PMI). In particular, comparing the performance of tradable and non-tradable sectors in the economies at the core of the Euro Area and in the periphery, Goldman Sachs' study

finds evidence of a re-orientation of activities from non-tradable to tradable productions in peripheral countries and from tradable to non-tradable in the core; results confirm that sizable sectorial shifts are taking place in national economies, the adjustment moving in the expected direction according to standard economic theory.

**(v) Micro-econometric methods**

10.38 Most models explaining aggregate outcomes, such as business cycles and inflation dynamics, include information about agents' opinions and expectations. Indeed, expectations are increasingly considered as a key factor in driving fluctuations in macroeconomic aggregates (see for instance, Leduc and Sill, 2013 for an analysis based on survey data). Nevertheless, how economic agents form their expectations about economy is less well studied or understood. In this respect, ETS may be of great help: indeed, survey questionnaires not only report about agents' opinions and expectations on important economic variables, but also allow to distinguish opinions according to individual socio-demographic (see for example on this issue Malgarini and Margani, 2008) and structural characteristics of the respondent (see chapter 3 for a list of the most common individual controls usually included in business and consumers tendency survey questionnaires). Information on opinions and expectations stemming from ETS which are often used in the literature may include inflation, the level of orders, demand and production, the business situation and access to credit for both households' and firms.

10.39 Many papers using micro level information have concentrated on CTS. Among the most prominent papers in this field, Carroll (2003 and 2006) used the US Michigan consumers survey to test the "epidemiological expectations" hypothesis, according to which households form their expectations by observing professional forecasts which are reported in the news media. However, according to this view households are supposed to observe the professional forecasts imperfectly, in the sense that initially only a subset of the population is capable of 'absorbing' the information, and only after some time the professional forecasts are eventually transmitted throughout the entire population; this hypothesis is found to be broadly consistent with the data. In a similar strand of literature, Easaw et al. (2013) use the ISTAT CTS to find that Italian household inflation expectations in the short run are excessively sensitive to their perception of current inflation rates, while in the long run they are anchored on professional forecasts. Results also appear to be highly policy relevant: in fact, while professional forecasters anchor their expectations on the ECB target, Italian households are found to settle on a considerably higher level, a discrepancy which has some important theoretical implications for policymakers. Departures from the Muthian rationality

hypothesis have also been investigated using survey data by Bovi (2009 and 2012): in his first contribution, Bovi uses a large dataset covering ten European countries for 22 years, finding that permanent and widespread psychological biases affect both the subjective probability of future economic events and their retrospective interpretation. Such biases can be explained on the basis of the psychological theory, according to which agents, when things go bad, tend to become particularly bullish, amplifying the forecast error; similarly, personal/future conditions are systematically perceived to be better than the aggregate/past ones. On similar grounds, Bovi (2012) finds that survey expectations are capable to enhance the forecasting ability of even the more statistical efficient econometric models, providing further contradiction to the standard hypothesis of rationality.

10.40 More recently, various authors have also used micro level information stemming from BTS. Rottman and Wolmershauer (2013) use data stemming from the monthly IFO Business survey on the manufacturing sector in order to derive an aggregate indicator of credit crunch for the German economy. Using the results of a monthly question on the current willingness of banks to extend credit to German firms, they estimate the probability of a restrictive loan supply policy by time. Estimations are conditioned on the creditworthiness of borrowers: the latter is approximated by firm-specific factors also derived from the survey, including firms' assessments of their current business situation and their business expectations. Using these data, the authors are also able to derive a credit crunch indicator, which measures the part of the shift in the loan supply that is neither explained by firm-specific factors nor by the opportunity costs of providing risky loans. Similar results for Italy are found by Costa et al. (2012) using the ISTAT BTS data; they are also able to extend the analysis to the services, retail and construction sectors and to use a richer set of controls for characterizing firms creditworthiness. A further example of a full exploitation of the information content of survey data comes from a paper by Basile et al (2012), focusing on business cycle asymmetries at the regional level: the authors find that firm-specific factors derived from BTS on the manufacturing sector such as borrowing constraints, export propensity, liquidity constraints and idiosyncratic demand shocks have an important role in explaining regional business cycle differentials in Italy. These factors are commonly neglected in more standard analysis typically focused on the role of the industry mix and the political cycle. Finally, survey data have also been used to study pricing strategies of exporting firms: Basile et al. (2014) find evidence of non-negligible reactions of export domestic price margins to unanticipated changes in cost competitiveness and in foreign and domestic demand levels, even though these effects appear to be of a transitory nature.

**References**

Abberger, K. and W. Wolfgang (2010). *The IFO business cycle clock: Circular correlation with the real GDP*. CESifo working paper: Empirical and Theoretical Methods, No. 3179.

Acemoglu, D. and A. Scott (1994). *Consumer confidence and rational expectations: Are agents beliefs consistent with the theory?* . The Economic Journal, 104, 1-19.

Aguiar-Conraria, L. and Soares M. J. (2011). *Business cycle synchronization and the Euro: A wavelet analysis* . Journal of Macroeconomics, 33(3), 477-489.

Al-Eyd, A., R. Barrell, R., and E. P. Davis (2009). *Consumer confidence indices and short-term forecasting of consumption.* Manchester School, 77(1), 96–111.

Angelini, E., G. Camba-Mendez, D. Giannone, L. Reichlin and G. Runstler (2011). *Short-term forecasts of Euro area GDP growth* . Econometrics Journal, 14, 25-44.

Artis, M., M. Marcellino M. and T. Proietti (2005). *Business Cycles in the New EU Member Countries and their Conformity with the Euro Area* . Journal of Business Cycle Measurement and Analysis, 2(1), 1-35.

Bachmann, R., S. Elstner, and E. R. Sims (2013). *Uncertainty and Economic Activity: Evidence from Business Survey Data.* American Economic Journal: Macroeconomics, 5, 217-49.

Baffigi, A., R. Golinelli, and G. Parigi (2004). *Bridge Models to Forecast the Euro Area GDP*. International Journal of Forecasting, 20 (3), 447-460

Basile B., S. de Nardis S. and A. Girardi, (2009). *Pricing to market of Italian exporting firms*. Applied Economics, vol. 41(12), 1543-1562.

Basile R., de Nardis and C. Pappalardo (2014). *Firm Heterogeneity and Regional Business Cycles Differentials*. Journal of Economic Geography, forthcoming

Bloom N. (2009). *The Impact of Uncertainty Shocks* Econometrica, 77 (3), 623-685.

Boivin, J. and S. Ng (2006). *Are more data always better for factor analysis?*.Journal of Econometrics, 132, 169-194Bovi M. (2009), Economic versus psychological forecasting. Evidence from consumer confidence survey, Journal of Economic Psychology, vol. 30

Bovi, M., (2013). *Are the representative agent's beliefs based on efficient econometric models?*. Journal of Economic Dynamics and Control, 37(3), 633-648.

Bruno G., (2014). *Consumer confidence and consumption forecast: a non-parametric approach*. Empirica, 41, 37–52.

Bruno G. and C.Lupi, (2004). *Forecasting industrial production and the early detection of turning points*. Empirical Economics 29, 647-671.

Burns A. F., W. Mitchell (1946)*Measuring Business Cycles*. NBER.

Carroll, C. D. (2003). *Macroeconomic expectations of households and professional forecasters*. The Quarterly Journal of Economics, 269-298.

Carroll, C. D. (2006). *The Epidemiology of macroeconomic expectations*. in The Economy as an Evolving Complex System, III, edited by L. Blume and S. Durlauf. Oxford: Oxford University Press.

*Centre for International Research on Economic Tendency Surveys*, www.ciret.org

Clemen, R.T. (1989). *Combining forecasts: a review and annotated bibliography*. International Journal of Forecasting, 5, 559-583

Costa, S., M. Malgarini and P. Margani (2012). *Access to credit for Italian firms: new evidence from the ISTAT confidence business surveys*. MPRA Paper 41389, University Library of Munich, Germany.

Croushore, D. (2005). *Do consumer-confidence indexes help forecast consumer spending in real time?*. The North American Journal of Economics and Finance, 16(3), 435-450.

Croushore, D. and T. Stark (2001). *A real-time data set for macroeconomists*. Journal of Econometrics, 105, 111-130.

De Boef, S., P.M. Kellstedt (2004). *The political (and economic) origins of consumer confidence*. American Journal of Political Science, 48, 633–649.

Dees S. And Soares Brinca P. (2013). *Consumer confidence as a predictor of consumption spending: Evidence for the United States and the Euro area*. International Economics, 134, 1-14.

De Leeuw, F. (1991). *Towards a theory of Leading Indicators* , in Lahiri and Moore (ed.), Leading Economic Indicators, New Approaches and Forecasting Records, Cambridge University Press, Cambridge.

Desroches, B and M. A. Gosselin (2004). *Evaluating Threshold Effects in Consumer Sentiment*. Southern Economic Journal, 70 (4), 942-952.

Diebold, F.X. and G. Rudebusch, (1996). *Measuring Business Cycles: A Modern Perspective* . Review of Economics and Statistics, 78, 67-77.

Diron M. (2008). *Short-term forecasts of euro area real gdp growth: An assessment of real-time performance based on vintage data*. Journal of Forecasting, 27(5), 371–390.

Easaw, J. R. Golinelli and M. Malgarini (2013). *What determines households inflation expectations? Theory and evidence from a households survey*. European Economic Review,61, 1-13.

ECB (2013). *Confidence Indicators and economic developments*. Monthly Bulletin, January. http://www.ecb.europa.eu/pub/pdf/mobu/mb201301en.pdf

Emerson, R.A. and D.F. Hendry (1996). *An evaluation of forecasting using leading indicators*. Journal of Forecasting, 15, 271-91.

Eurocoin. http://eurocoin.cepr.org/

European Commission, Business and Consumers Surveys Portal, http://ec.europa.eu/economy_finance/db_indicators/surveys/index_en.htm

European Commission, Directorate-General for Economic and Financial Affairs (2014), *"The Joint Harmonised EU Programme of Business and Consumer Surveys"*, user guide March, available at http://ec.europa.eu/economy_finance/db_indicators/surveys/method_guides/index_en.htm.

EUROSTAT (2010). *The European Business Cycle Clock*.http://epp.eurostat.ec.europa.eu/cache/BCC/explanation_en.html

Farmer R. E. A. (1999). *Macroeconomics of Self-fulfilling Prophecies* . 2nd Edition, MIT Press Books, The MIT Press, edition 2, vol. 1, August.

Forni M., M. Halli, M. Lippi and L. Reichlin (2005). *The Generalized factor model: One-sided estimation and forecasting*. Journal of the American Statistical Association, American Statistical Association, 100, 830-840.

Gayer, C. (2007). *A fresh look at business cycle synchronization in the Euro Area*. European Economy, Economic Papers, n. 287, Brussels.

Gayer, C. (2010). *Report: The Economic Climate Tracer-A Tool to Visualise the Cyclical Stance of the Economy Using Survey Data*. EU Commission, Brussels.

Gayer, C. and J. Genet (2006). *Using factor models to construct composite indicators from BCS data – A comparison with European Commission confidence indicators*. European Economy, Economic Papers 240, European Commission, Brussels

Gayer C. and P. Weiss (2006). *Convergence of Business Cycles in the Euro Area: Evidence from Survey Data*. Applied Economics Quarterly, 52(3),

Garner, C. A. (2002). *Consumer confidence after September 11*.Federal Reserve Bank of Kansas City, Economic Review 2, 1–21.

*Giannone D., L. Reichlin, S. Simonelli (2009). Nowcasting Euro Area Economic Activity In Real Time: The Role Of Confidence Indicators*. National Institute Economic Review, National Institute of Economic and Social Research, 210(1), 90-97.

Goldman Sachs (2013*). Godman Sachs Newsletter*. July.

Golinelli, R. and G. Parigi (2004). *Consumer sentiment and economic activity: A cross country comparison*. Journal of Business Cycle Measurement and Analysis, 1, 147-170.

Golinelli, R., G. Parigi (2005). *Short-Run Italian GDP Forecasting and Real-Time Data*. CEPR Discussion Papers 5302, C.E.P.R. Discussion Papers.

Golinelli, R. and G. Parigi (2007). *The use of monthly indicators to forecast quarterly GDP in the short run: an application to the G7 countries*. Journal of Forecasting, 26 (2), 77-94.

Graff M. and J. E. Sturm (2012). *The Information Content of Capacity Utilisation Rates for Output Gap Estimates*. CESifo Economic Studies, 58 (1), 220–251.

Hamilton J. (1990). *Analysis of Time Series Subject to Changes in Regime*. Journal of Econometrics, 45(1-2), 39-70.

Harrison S. G. and M. Weder (2006). *Did sunspot forces cause the Great Depression?*. Journal of Monetary Economics, Elsevier, 53(7), 1327-1339.

Howrey E. P. (2001). *The Predictive Power of the Index of Consumer Sentiment*. Brookings Papers on Economic Activity, The Brookings Institution, 32(1), 175-216.

Katona, G. (1975). *Psychological Economics*. Elsevier Scientific Publishing Company, New York.

Kim S. and S. Kim (2013). *International capital flows, boom-bust cycles, and business cycle synchronization in the Asia Pacific region*. Contemporary Economic Policy, 31(1), 191-211.

Kose M. A., C. Otrok and E. Prasad (2012). *Global Business Cycles: Convergence Or Decoupling?*. International Economic Review, 53(2), 511-538.

Koopmans T. (1947). *Measurement without theory*. Review of Economic Statistics 29 (3), 161–72.

IFO Business Cycle Clock. http://www.cesifo-group.de/ifoHome/facts/Survey-Results/Business-Climate.html

INSEE Composite indicators. http://www.insee.fr/en/themes/indicateur.asp?id=105

Leduc S. And K. Sill (2013). *Expectations and Economic Fluctuations: An Analysis Using Survey Data*. The Review of Economics and Statistics, 95(4), 1352-1367.

Lovell, M. C. (1975). *Why was the consumer feeling so sad?*. Brooking Papers on Economic Activity, 2, 473-479.

Ludvigson, S. C. (2004). *Consumer confidence and consumer spending*. Journal of Economic Perspectives, 18(2), 29–50.

Malgarini, M. and P. Margani (2007). *Psychology, consumer sentiment and household expenditures*. Applied Economics, 39(13), 1719–1729.

Malgarini, M. and P. Margani (2008). *Unemployment expectations across heterogeneous groups*. Economics Bulletin, 4 (31),1-6.

Malgarini, M., M. Mancini and L. Pacelli, (2013). *Temporary hires and innovative investments*. Applied Economics, 45(17), 2361-2370.

Matsusaka, J. G. and A. M. Sbordone (1995). *Consumer confidence and economic fluctuations*. Economic Inquiry, 33(2), 296–318.

OECD (2003). *Business Tendency Surveys: A handbook.* OECD Publications, Paris, ISBN 92-64-19894-6. (http://www.oecd.org/std/leading-indicators/31837055.pdf.

OECD. *Leading Indicators and Tendency Surveys Portal*. http://www.oecd.org/std/clits/

OECD. *System of Composite Leading Indicators*, http://www.oecd.org/std/leading-indicators/

OECD (2010). *Business Cycle Clock*. http://stats.oecd.org/mei/bcc/default.html

Ramalho, E. A., A. Caleiro and A. Dionfsio (2011). *Explaining consumer confidence in Portugal*. Journal of Economic Psychology, 32, 25-32.

Rottman H. and T. Wollmershauser (2013). *A micro data approach to the identification of credit crunches*. Applied Economics, 47 (17), 2423-2441.

Roos, M. W. (2008). *Willingness to consume and ability to consume*. Journal of Economic Behavior & Organization, 66, 387–402.

Stock J. M. and M. W. Watson (1999). *A Comparison of Linear and Non-Linear Univariate Models for Forecasting Macroeconomic Time Series*. in: R. Engle and H. White (eds.) Cointegration, Causality and Forecasting: A Festschrift in Honor of Clive W.J. Granger, Oxford University Press, Oxford.

Stock J.M. and M.W. Watson (2002a). *Macroeconomic forecasting using diffusion indexes.* Journal of Business & Economic Statistics, 20, 147-161.

Stock J.M. and M.W. Watson (2002b). *Forecasting using principal components from large number of predictors.* Journal of the American Statistical Association, 97, 1167-1179.

Ruth, V. and B. Schouten and R. Wekker (2005). *The Statistics Netherlands' Business Cycle Tracer. Methodological aspects; concept, cycle computation and indicator selection.* Statistics Netherlands report 2005-MIC-44.

Taylor K. and R. McNabb (2007). *Business Cycles and the Role of Confidence: Evidence for Europe.* Oxford Bulletin of Economics and Statistics, 69(2), 185-208.

Throop, A. W. (1992). *Consumer sentiment: its causes and effects*. Economic Review, 35–59.

Vuchelen, J. (1995). *Political events and consumer confidence in Belgium*. Journal of Economic Psychology, 16, 563–579.

Zarnowitz, V. (1992*). Business Cycles, Theory, History, Indicators, and Forecasting.* The University of Chicago Press, Chicago.