ST/ESA/STAT/SER.M/96

Statistical papers Series M No. 96

Handbook on

Economic Tendency Surveys

UNITED NATIONS

New York, 2015

Note

The designations used and the presentation of material in this publication do not imply the expression of any opinion whatsoever on the part of the Secretariat of the United Nations concerning the legal status of any country, territory, city or area, or of its authorities, or concerning the delimitation of its frontiers or boundaries.

The term "country" as used in this publication also refers, as appropriate, to territories or areas.

The designations "developed regions" and "developing regions" are intended for statistical convenience and do not necessarily express a judgment about the stage reached by a particular country or area in the development process.

Symbols of United Nations documents are composed of capital letters combined with figures. Mention of such a symbol indicates a reference to a United Nations document.

ST/ESA/STAT/SER.M/96

UNITED NATIONS PUBLICATION
Sales No.: E.16.XVII.7
ISBN: 978-92-1-161604-0

Acknowledgements

The preparation of *the Economic Tendency Survey Handbook* was initiated by the Statistical Commission as part of the international programme of work on short-term economic statistics which was developed in response to the 2007/2008 economic and financial crisis. The international programme on short-term economic statistics was the result of a wide consultation - initiated by the United Nations Statistics Division and Statistical Office of the European Communities (Eurostat) in collaboration with Statistics Canada, Statistics Netherlands and the Russian Federal State Statistics Service - in order to ensure a coordinated statistical response to the economic and financial crisis. Four themes were identified in the programme: business cycle composite indicators, economic tendency surveys, rapid estimates, and data template and analytical indicators.

For each thematic area, the Commission approved the preparation of handbooks with a view to providing guidance, best practices and harmonized principles to assist member States in compiling and reporting internationally comparable short-term statistics.

The Economic Tendency Survey Handbook aims at providing best practices and harmonized principles on tendency survey sample selection, questionnaire design, survey questions, survey execution, data processing, dissemination of results and use of composite tendency indicators.

The Handbook was drafted by a working group comprising the Italian National Institute of Statistics (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) and the United Nations Statistics Division. The working group was led by ISTAT. The European Commission (EC) and

the French National Institute of Statistics and Economic Studies (INSEE) have also contributed to the review of the draft chapters of the handbook.

The handbook benefited from the contributions of many experts, most notably: Mauro Politi, Luciana Crosilla, Roberto Gismondi, 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); Aloisio Campelo Junior and Viviane Seda Bittencourt (Instituto Brasileiro de Economia, Fundação Getulio Vargas, IBRE-FGV); Christian Gayer, Roberta Friz, Andreas Reuter, Alessandro Girardi, (EC - DG ECFIN); Cedric Audenis (INSEE); Gulab Singh, Ilaria Di Matteo and Herman Smith (UNSD).

The Handbook has also benefited greatly from the numerous useful comments and suggestions made by national statistical offices, central banks, regional commissions, and international organizations, as well as individual experts, during the global consultation which took place from July to September 2014.

The invaluable support to the coordination and finalization of the Handbook by Mauro Politi, Luciana Crosilla and Jolanda Giaconi (ISTAT) is acknowledged with gratitude.

This publication was prepared under the responsibility of the Statistics Division under the guidance of Herman Smith and Ivo Havinga.

Table of Contents

BOX

FIGURES

TABLES

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 Surveys
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 Surveys
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 Surveys
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	Partitioning 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
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
UNSD	United Nations Statistics Division
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/consumers about the current situation of their business/household, their opinions on the whole economy and their plans and expectations for the near future. Generally, respondents are not asked for quantitative information such as, for example, levels of output, sales, investment, employment or household finance; respondents instead are asked to provide qualitative information.

1.2 The first ETS originated in the 1920's, and they have been implemented by an increasing number of countries over the years. The increasing importance of ETS can be seen by the work that has been carried out at regional and international level. In Europe, for example, 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 (i.e. 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 (EC DG ECFIN, 2014). The Organisation for Economic Co-operation and Development (OECD) also is active with tendency surveys: OECD has collected best practices and compiled a handbook in 2003 (OECD, 2003). A lot of advancement and progress in tendency surveys was also due to the work of the Centre for International Research

[1] Since 2007, the Commission conducts survey also on Financial and insurance activities at EU and euro-area level.

on Economic Tendency Surveys (CIRET) which has been active since 1953 in advancing research and development in ETS. CIRET gathers compilers and users of ETS to promote the exchange of theoretical and operational knowledge on economic cycle research.

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

1.4 ETS provide valuable insight and information both for countries with well-established and advanced statistical system and also for countries with a weaker statistical system. In the first case, ETS provides valuable information additional to the existing set of quantitative statistics; in the second case, ETS are often one of a few - or even the only - source of high frequency economic data.

1.5 After the 2007-2008 financial crises, the international statistical community gathered together to discuss the role of official statistics to provide tools for policymakers to monitor trends, measure impacts and provide early warnings in order to formulate informed policies to minimize the impacts of such events. An extensive discussion and consultation was carried out among international/regional organizations and countries to identify and remedy data gaps for monitoring economic shocks as well as to improve the dissemination and communication of relevant and already available information. The outcome of this consultation formed the basis for the formulation of an international programme of work on short-term economic statistics as part of a coordinated statistical response to the economic and financial crisis.

1.6 The international programme of work on short-term economic statistics was prepared by the United Nations Statistics Division (UNSD) and Eurostat and was endorsed by the United

Nations Statistical Commission in 2011[2]. In this programme, four thematic areas of work were identified: business cycle composite indicators; economic tendency surveys; rapid estimates; and a data template and analytical indicators. Tendency surveys were recognised as useful tools in providing an early warning of changes in economic activity either as stand-alone indicators or as components of coherent indicator sets and composite indicators. They are thus an important part of the statistical system.

1.7 This handbook provides internationally harmonized principles for the conduct of tendency surveys. It builds on the existing work done by countries and international/regional organizations, most notably the work done as part of the Joint Harmonised EU Programme of Business and Consumer Surveys and its guidelines (EC DG ECFIN, 2014) and the work done by the OECD (OECD, 2003). This handbook covers both types of economic tendency surveys: business tendency surveys (BTS) and consumer tendency surveys (CTS) and introduces flexibility for countries to adapt ETS to the specific country's context.

1.8 The recommendations provided in this handbook are applicable to all countries irrespective of the development of their statistical systems and are important for international comparability. It is important to harmonize practices through the use of a common set of questions and methods in ETS to ensure the quality and reliability of survey results and to facilitate the international comparison of the ETS results. This will result in an increased value and trust in ETS at the national and international levels.

1.9 It should be mentioned, however, that in conducting ETS there is some degree of flexibility to adapt to national circumstances. This means that there could be particular types of economic activities that are important to survey for monitoring the national economy, which are

[2] See Report of the UNSC at its forty-second session document E/2011/24-E/CN.3/2011/37, available at: http://unstats.un.org/unsd/statcom/doc11/Report-Final-E.pdf

not covered by other countries; or there could be specific questions that can be added to complement the set of internationally agreed questions. This might be justified by specific market conditions or by institutional, legal or infrastructure peculiarities. Even with regard to the survey process, the handbook provides a review of sampling techniques that can be used in ETS; the actual sampling procedure, the survey mode, and the survey compilation methods have to be chosen taking into account the specific national conditions and based on sound methodological considerations.

1.10 The target audience of this handbook consists not only of compilers of ETS that are in the stage of setting up or strengthening the programme for tendency surveys, but also users of ETS seeking to gain a better understanding of the methodological underpinning of ETS data. There is a wide variety of ETS users: analytical users that use ETS, for example, to compile composite indicators; researchers, policymakers, academia and the media who use ETS and derived indicators, such as sentiment indicators and other composite indicators, that are interested in monitoring and analysing economic conditions and business cycles.

1.11 It should be noted that the institutional set-up for ETS differs among countries according to their specific situation. ETS are conducted by National Statistical Offices, private or public institutes. For instance, in the United States, the most prominent ETS are conducted and released by private institutes—for example, the consumer confidence index from the Conference Board, the index of consumer sentiment from Thompson Reuters and University of Michigan, and the Purchasing Managers Index (PMI) from the Institute for Supply Management. In India, the Central Bank is active in business tendency surveys. In Europe a number of National Statistical Offices compile ETS (for example, INSEE in France and ISTAT in Italy). The

recommendations provided in this handbook remain valid irrespective of the type of institutional set-up for the compilation of ETS in countries.

B. Characteristics of Economic Tendency Surveys

1.12 ETS collect *qualitative data* on a wide range of topics. Entrepreneurs and consumers are asked about 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-defined answers, describing the past, current and future developments in qualitative statements. This practice is rather different from that of quantitative surveys, which ask for a numeric evaluation of the variable of interest.

1.13 ETS have some unique characteristics. The most important feature of ETS is their *timeliness*. The simple and intuitive questionnaire makes it very easy for the respondent to complete it quickly. There is no need for the respondent to obtain exact figures from within his/her company/households, since the survey asks for the respondent's indicative statements on developments and expectations. Because of this simplicity of the questionnaire, the processing time of the data is also very short: usually results of ETS are available soon after the end of the reference period.

1.14 Another important characteristic of ETS is that they provide a unique source of information about agents' expectations on relevant economic outcomes (e.g. production, demand, households' income), and they allow for the collection of information on, for example, constraints on production which are generally not otherwise collected, and/or inventories of behaviour for which quantitative data are generally insufficient. Furthermore, since tendency

surveys are generally not subject to revisions (see more in Chapter 9) and their results don't show a trend[3], they are particularly useful for business cycle analysis.

1.15 Tendency surveys are not a substitute for quantitative data, but they complement quantitative data. Since ETS collect judgments and expectations for the past and the near future, they are able to detect changes in the cycle earlier than the corresponding quantitative statistical series; this is because judgments and expectations lead to plans and only when these plans are implemented are they picked up by traditional quantitative statistics. When quantitative series are available, this link may be verified with consistency checks provided by the comparison of ETS data with corresponding relevant quantitative series such as, GDP, consumption, employment, etc. (additional information on this issue is provided in Chapter 10). A description of the theoretical debate regarding consumer tendency surveys and their links with economic behaviour is described in Box 1.1.

Box 1.1: Theoretical debate on consumer tendency surveys

Additional considerations have to be made to better understand the importance of consumer tendency surveys. 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.

The original insight behind the setting of CTS is due to George Katona (1951), who 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

[3] This feature depends on the method of quantification of the answers (for further details see chapter 8).

also on the "willingness to buy". Thus, 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 crises or wars – that are 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 to turning points.

Katona concludes that people's expenditures are not only a function of income, assets, age and other objective factors. It is probable that attitudes are to some extent important. The wealthier a community is, the more likely it is for "subjective" factors to influence spending behaviour. Usually a substantial portion of people's income is absorbed by routine purchases and people save what is left over. With rising income, people gain the ability to spend or save. So consumer surveys become increasingly important the more an economy develops.

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. A first class of models attributes only an indirect function to sentiment, which is considered merely as a predictor of current income, in case of failure of some restrictive hypothesis of the Permanent Income Hypothesis (PIH), namely capital market efficiency and absence of uncertainty (Carroll et al., 1994; Campbell and Mankiw, 1991; Acemoglu and Scott, 1994).

According to a different view, sentiment has a direct impact on consumption, reflecting the role of habits (Deaton, 1992; Sommer, 2001): agents may become addicted to the level of consumption experienced in the past and adjust their expenditure pattern only gradually, responding late to news on the general and personal economic situation.

Another theory links consumers' sentiment to precautionary motives (Carroll, 1992): an increase in uncertainty – as measured by a decline of consumers' sentiment index (CSI) - may cause a higher perception of the probability of financial distress, leading consumers to save more in liquid assets and less in illiquid assets, postponing expenditures on durables rather than non-durables goods and services.

Expectations about the future level of output can become self-fulfilling (Matsusaka and Sbordone, 1995), with the result that a decline in sentiment can cause a fall in output and consumption growth, even after controlling for economic fundamentals. To some extent, this last theoretical justification is related to Katona's psychological approach, in the sense that the unwillingness to buy may be interpreted as a case of coordination failure.

Recent literature has pointed to a possible role for group-specific shocks that may hit differently different group of consumers. This may be due to the fact that households pay attention in a different way to macroeconomic news, depending on their level of income, education or in general their socio-demographic characteristics (Carroll, 2004).

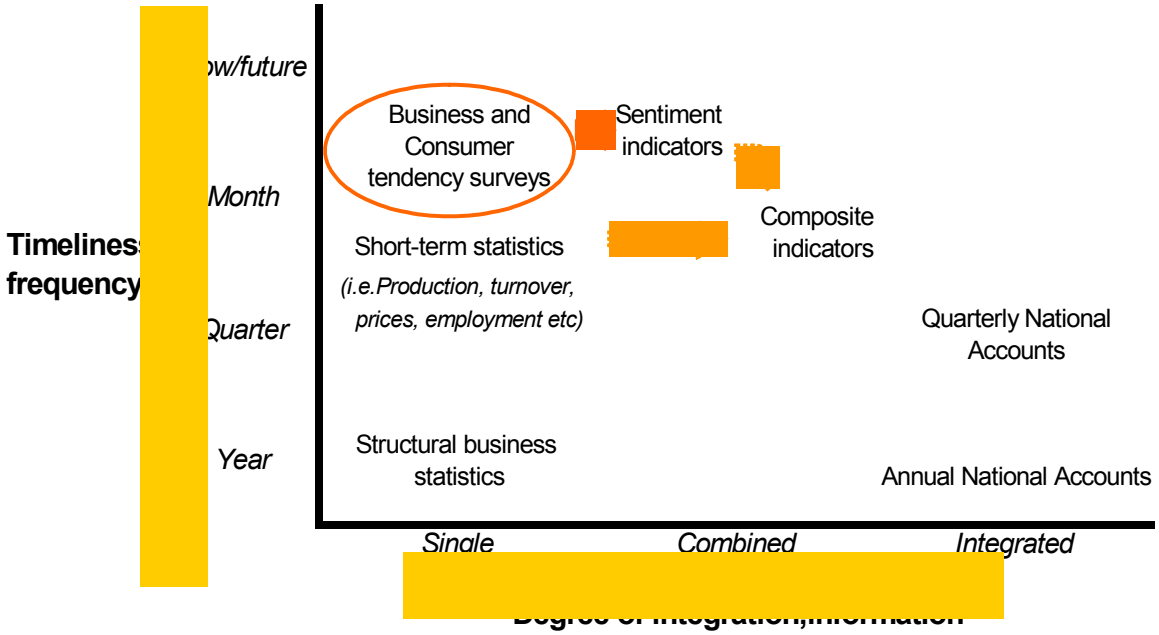
This view is shared by the view of Souleles (2004), which analyses the consumption-sentiment relationship, exploiting the micro-level nature of the data and controlling for household demographic characteristics: he finds that US CSI effectively helps to forecast consumption dynamics, in contrast with the PIH. In this case, excess sensitivity of consumption expenditures to sentiment proves to be partly explained by the inclusion of social-demographic components, indicating an important role for group-specific shocks, consistent with skill-biased technical change.

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 sceptical position taken by the FED (1955), according to which sentiment can be reasonably well approximated by a set of standard macroeconomic variables, having no impact on consumption patterns.

1.16 In order to better understand how tendency surveys fit into the overall system of economic statistics, Figure 1.1 shows how different types of economic statistics are related to each other, taking into consideration frequency and timeliness as well the degree of integration of the information.

Figure 1.1: ETS and the system of economic statistics



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

1.17 The degree of integration of the statistics is represented in the figure in the horizontal axis. The label "single" refers to the statistics obtained from a single survey of statistical units.

The data are simply the survey findings. The label "combined" refers to the statistics that are the outcome of combining different statistics (and different surveys). They are the result of taking statistics together, but are not yet integrated. An example is the deflation of turnover (from one source) using prices (from another source e.g. price statistics) and weighting schemes (from another source again e.g., national accounts). Lastly, the label "integrated" contains data that are the outcome of an integration process involving detailed checks of all available information. The national accounts and the quarterly national accounts are examples of integrated statistics.

1.18 The vertical axis refers to the time reference of the data - monthly, quarterly and annual data - which is also strongly related to their timeliness. Reading the figure from top to bottom on the vertical axis, the reported results become more reliable and detailed, but less timely. Reading the figure from left to right, the data become more comprehensive and reliable, but again as a general rule less timely. Economic tendency surveys are compiled on a monthly basis and the data are available shortly after the reference period. Other short-term statistics (such as production, turnover, employment etc.) while compiled on a monthly basis as well, are usually released within a few weeks after the end of the reference period. Structural business statistics are compiled on an annual basis and are usually available a year from the end of the reference period.

1.19 The ETS are uniquely located within the system as data that are quickly available - either before of the end of the reference period or very shortly after the end of the reference period - and they also provide information about expectations for the near future.

C. Uses of Economic Tendency Surveys

1.20 ETS data are used by a variety of users' groups. Generally, all those who are involved in business cycle analysis are potential customers of ETS data. Politicians, ministries, central banks,

commercial banks, financial market analysts, and professional associations are also users of ETS results (for further details about users of survey data, see Chapter 10). In addition, since the results of ETS are usually fairly simple to understand and have a predictive element, they are appealing also to ordinary people who like to see what the future holds and thus they often attract media reports.

1.21 Businesses themselves are often interested in business tendency surveys (BTS) results at an industry level. The information helps them to compare their own position with the industry average. Detailed data about customer sectors or supplier sectors can help to optimize their own production processes. There is even interest in the survey data at the micro-level: scientists use anonymized micro-data for research.

1.22 ETS results are used in several ways. They are used for the compilation of sentiment indicators but also, by combining them with other short-term statistics, for constructing business cycle composite indicators with various methodologies[4]; in addition, this information can be used to calculate "confidence" indicators by economic activity, for the whole economy and for consumers. A well-established indicators system, which uses data from various sources including ETS, is the system of Composite Leading Indicators (CLIs) of the OECD.

1.23 ETS data are also used as visualizing tools of the business cycle situation and as input variables in short-term nowcasting and forecasting models. In addition, ETS results are included in macroeconomic analysis and as a source of information for micro-econometric analysis of agents' behaviour (see Chapter 10).

1.24 The heterogeneity of the users' groups and of the uses of ETS means that the institution conducting and compiling ETS should develop a well-thought-out publication strategy, which

[4] A useful guide for constructing composite indicators is the forthcoming UN Handbook on Cyclical Composite Indicators.

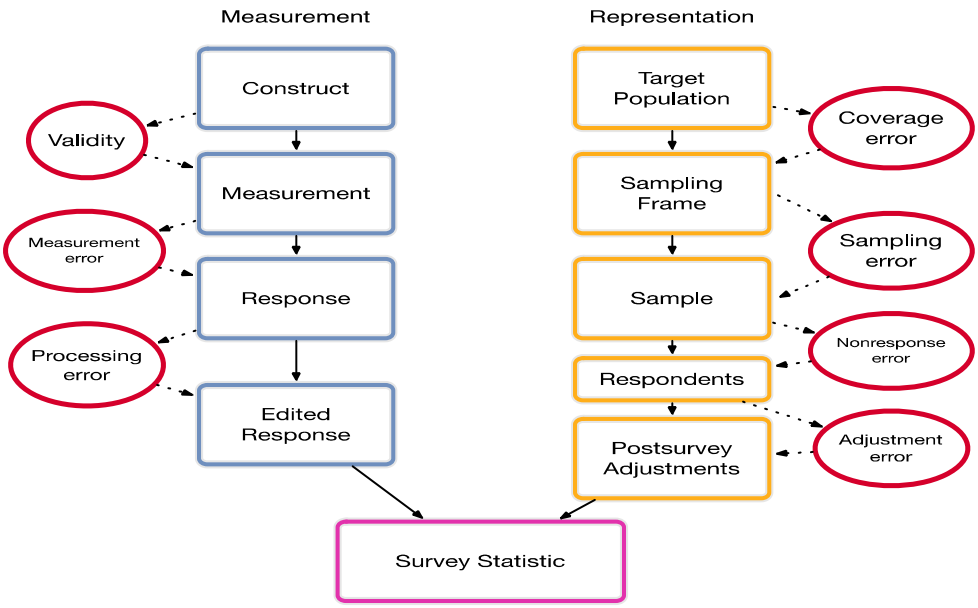
addresses the needs of the different user groups. Media are often interested in brief and timely press releases. Policymakers need more in-depth reports and businesses are interested in results by industries. In addition, if micro-data are released to particular user group, attention should be given to guaranteeing the confidentiality of the information.

D. Survey process

1.25 As the name suggests, ETS are based on sample surveys and thus they are subject to the usual survey errors. This section aims at showing how this handbook helps to minimize some of these errors. Figure 1.2, in particular, shows common sources of errors associated with the measurement aspect and the representation aspect of a survey. The measurement aspect of a survey deals with the question: what is this survey about? The representational dimension focuses on: who is the survey about?

Figure 1.2: Possible sources of errors in surveys

Source: Groves et al., (2009)



1.26 ETS aims at capturing the economic situation of a country, usually the business cycle development. In order to measure the economic situation, various questions were developed and over time a core set of questions, which are used successfully in many countries, has emerged. The questions and the pre-defined response categories are chosen in such a way that respondents are able to answer relatively quickly and on the spot. In general, no additional information or calculation is necessary for answering the questions. These aspects distinguish ETS from other surveys (e.g. quantitative surveys). Since the questions and the response scales proposed in this handbook have already been successfully used in many countries for a long time, errors in the measurement process should be minimal. Sometimes there might be reasons to adapt the standard approach because of cultural, social or other reasons. These adaptations should try to reduce errors in the measurement process.

1.27 Recommendations are made in this handbook with regard to the representation dimension of the ETS. The handbook addresses questions such as: 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 frame 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 practices for BTS and CTS process are described which also allow adaptations to local circumstances.

E. Structure of the handbook

1.28 This handbook covers the whole statistical production process for ETS in all its aspects. As such it addresses issues in the planning of the survey, collection, processing and dissemination of ETS results. It also provides examples of uses of the survey results.

1.29 Chapter 1 provides the introduction to the handbook. It defines the objectives and the structure of the handbook and introduces the main characteristics of the ETS.

1.30 Chapter 2 describes the objectives of ETS and contains the main characteristics of business and consumer tendency surveys in terms of the target populations, statistical units, standard classifications and survey frequency.

1.31 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. The pre-testing of the questionnaire is illustrated, too.

1.32 Chapter 4 focuses on important steps in the entire process of ETS, starting with where and how to select a subset of units to sample; survey frame and sample designs suitable for ETS are also illustrated. Moreover, the methods of selecting units from the survey frame, the allocation criteria and the sample size are discussed.

1.33 Chapter 5 presents considerations regarding estimation procedure and accuracy. In BTS and CTS estimation processes, weighting schemes are usually used. Referring to 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. The chapter also illustrates how to calculate sampling error.

1.34 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.

1.35 In Chapter 7 non-sampling errors and the possible approaches to reducing errors are analysed; in particular coverage, measurement, processing and non-response (total and partial) errors are discussed.

1.36 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.

1.37 Data dissemination and publication are the topics of Chapter 9. Recommendations regarding metadata and output tables are made; strategies of publication are also discussed.

1.38 Chapter 10 elaborates on the 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.

1.39 This handbook contains two Annexes. Annex 1 provides a detailed description of the coverage of the economic activities used in BTS in terms of the International Standard Industrial Classification of All Economic Activities, (ISIC Rev.4); Annex 2 provides the list of questions used in BTS and CTS on a monthly, quarterly and semi-annual basis.

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 "what" needs to be measured, i.e. what is the data need, and "who" data are asked from. In the case of BTS, one has also to determine what level of the business structure data are collected from, namely enterprises, establishments etc. and who, within the business structure, should respond to the survey. For CTS, one needs to determine if it is the individual consumer or the household that should be targeted. Other elements that determine the scope of ETS include: kinds of economic activities, geographical coverage, variables to be inquired, type of questions, and frequency of the survey. Table 2.1 provides an overview of the questions that should be addressed when constructing a survey.

Table 2.1: Scope of ETS

Who	Business or consumers? Which economic activities to survey? Enterprise or establishment? Individual consumer or households? Fixed sample or rotating sample?	**Sample design** (chapter 4)
Where	Geographical coverage?	
What	What data? What variables covered? What elements of the economy?	**Questionnaire design** (chapter 3)
How	Structure of the answering scheme?	
When	Frequency of the survey? Monthly, quarterly and/or semi-annual questions? Timing for data collection?	

2.2 ETS can be conducted among businesses or consumers. Both business and consumer tendency surveys can be used to analyse economic activity. Businesses give information on the supply side of the economy. Consumers give information from a demand perspective. The first decision in conducting ETS is to determine whether to conduct a business or consumer tendency survey (*Who*). Once this is decided, a further step is to refine the choices: in the case of BTS it means deciding which economic activities and firms are to be surveyed; and in the case of CTS which types of consumers/households are to be covered in the survey and what kind of panel (e.g. fixed or rotating panel) will be used.

2.3 Geographical considerations may play an important role in the planning for ETS (*Where*). 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 geographical differences are considered important, they should be addressed when designing the sample survey for ETS.

2.4 The third element is a selection of desired variables (*What*). Which questions are needed to give insight into economic developments? What target indicators have to be monitored? The questions can be backward- or forward-looking, that is, asking for assessments of the past situation or expectations for the near future. When designing the questionnaire, there should be a balance between the number of variables asked about and the burden on respondents in completing the questionnaire; and harmonization with international standards and country specific conditions. In addition, in order to make it possible to collect data in a very short period, the answering scheme must be easy and intuitive. Thus the questionnaires normally have a fixed number of pre-defined qualitative answering schemes even tough, in some cases, a quantitative indication may be required (*How*).

2.5 Finally, the frequency of the survey must be determined (*When*). It is recommended that ETS be conducted on a monthly basis. Some additional questions may be added on a quarterly and/or semi-annual basis. It is important to establish a well-organized data collection strategy in order to ensure a timely release of data (for further details see Chapter 3).

2.6 All these aspects are presented in more detail in the next sections for both business and tendency surveys.

B. The scope of Business Tendency Surveys

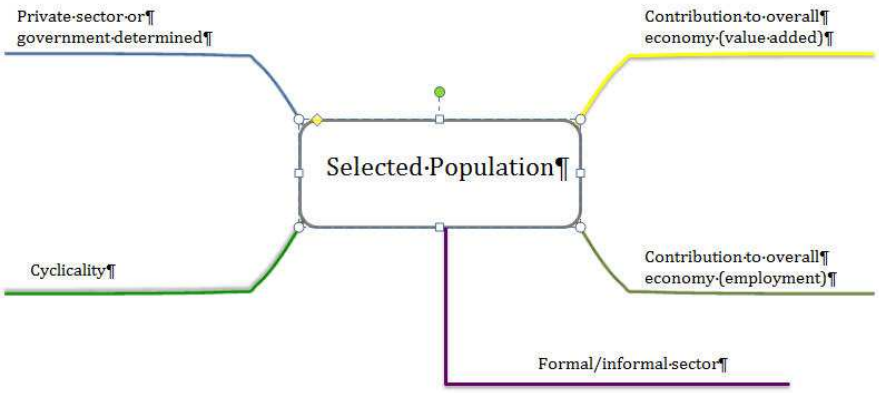
1. Kind of economic activities: the coverage

2.7 BTS are carried out to analyse developments in a specific segment of the economy and to obtain indicators for the situation in the economy as a whole.

2.8 To get a good overview of the state of the economy, the coverage of the BTS - in terms of economic activities - has to be chosen with care. Since BTS aim to give timely signals about the cyclical development of an economy, not all economic activities are surveyed by BTS.

2.9 When constructing a BTS, the choice of which economic activities to cover is an essential step. Figure 2.1 indicates the main factors that influence the selection of which economic activities to cover in BTS.

Figure 2.1: Factors relevant for the selection of kinds of economic activities



2.10 The *size* of a particular economic activity, measured for example by value added or the number of employed workers, in combination with its cyclicality, are important decision criteria. The *size* measures the contribution of the economic activity to the overall economy, while *cyclicality* refers to its sensitivity to changes in the business cycle. In other words, an economic activity is "cyclical" if it is sensitive to the business cycle, that is, for example, the revenues are generally higher in periods of economic prosperity and expansion and lower in periods of economic downturn and contraction. In general, since the goal of ETS is to give timely signals about the cyclical development of an economy, cyclicality is a fundamental feature and plays a role not only in the selection of activities to be surveyed but also in the measurement of the quality of the indicators. Some kinds of economic activities do not show much variation over the business cycle. These economic activities stabilize the economy but they do not provide strong signals about the business cycle.

2.11 Another criterion is whether the economic activities are controlled by the government or are private or semi-private corporations. Economic activities where the activity is determined by the government are often not of interest when the aim is to get timely signals about the cyclical development of an economy. Some kinds of economic activities move rather anti-cyclically, and

this can be disrupting especially when cumulating different surveys into a country-wide sentiment indicator.

2.12 Other characteristics that may be important to the decision on the coverage of specific segments of the economy are linked to: the likelihood of high response rates and reliable responses; the prevalence of micro-firms in the specific segment of the economy as well as the prevalence of informal activities[1]. Weather and seasonal influences must also be considered in selecting the kind of economic activities because economic activities influenced by these two factors tend to respond slowly, if at all, to movements in the business cycle.

2.13 Economic activities are classified according to the International Standard Industrial Classification of All Economic Activities, (ISIC Rev. 4). Annex 1 provides a detailed correspondence between the kind of economic activities generally covered in BTS and the relevant Sections and Divisions of ISIC Rev 4. Core kinds of economic activities, which are covered by BTS in many countries, are: Manufacturing (ISIC rev. 4 Section C); Construction (ISIC rev. 4 Section F); Retail trade (ISIC rev. 4, Divisions 45 (part) and 47); and Services (ISIC rev. 4, Sections H to N and Sections R-S).

2.14 Manufacturing is one of the most important industries to be covered by BTS. Some industry branches are very cyclical and, with the expenditure approach in national accounts in mind, the demand for machinery and equipment is an important cyclical component of fixed investment. Fluctuations in machinery and equipment expenditures are closely related to the manufacturing industry.

2.15 Another important indicator of fixed investment is construction investment. Thus it is very reasonable to cover also construction activities (ISIC rev. 4 Section F). In many emerging

[1] For additional information on data collection on the informal sector please refer to UNSD (2005) and ILO (2013).

countries construction activities are an important driver of economic development. In addition, these economic activities have often been involved in economic crises in different countries, emerging or industrialized.

2.16 Retail trade is usually covered because retailing is a quite volatile part of consumption and in many countries consumption contributes 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 to CTS should give a timely and reliable picture about cyclical movements of consumption.

2.17 Although manufacturing is often regarded as one of the most important contributors to the business cycle, the value added of Service activities is in many countries much larger than that of manufacturing activities. The types of economic activities included in Services are quite heterogeneous: there are services to businesses, services to individuals and services to government. 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 to a greater or lesser extent by the government (e.g. health services). Some services 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).

2.18 This heterogeneity is the reason why often only a selection of services is covered by BTS. The selection is more country-specific than for manufacturing, construction and retail trade as 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 about.

2.19 Another problem area within the services sections is the treatment of financial services. Although very important for most developed countries, there are some objections against including them in the survey for services. The number of financial services providers in a country is often limited, and also a part of the selected variables are not applicable to financial services. These drawbacks can be addressed by conducting a special survey among financial services providers.

2.20 Wholesale has various distribution functions in an economy. It distributes intermediate products to manufacturers and consumer goods to retailers. So wholesale is connected to both production and consumption activities. There is also another function of wholesale: import and export of goods. So wholesaling includes a reasonable choice of economic activity to cover in countries with large export or import activities. Therefore, this industry could be of special interest, for example, in developing countries with small manufacturing industries.

2.21 The inclusion of additional economic activities, such as Agriculture or Mining, can also be considered as country-specific. This decision should be guided by the factors presented in Figure 2.1, which include, among others, characteristics of size, cyclicality and market structure. However, they should be included only when they are at least important in terms of contribution to GDP. Agriculture is particularly influenced by weather conditions, seasonality and anti-cyclical behaviour. If it does not contribute substantially to GDP, its coverage might not be necessary for the assessment of the general economic situation.

2.22 Once the kind of economic activities is selected, the level of detail for publication must be determined. Common strata in BTS are type of activity (usually based on a two-digit level of detail of ISIC Rev. 4) and firm size (often measured by the number of employees or turnover). In some countries, the geographical dimension is covered as well (see below).

2. Geographical coverage

2.23 Defining the scope of BTS in terms of which economic activities to cover is an essential step in the set-up of such surveys. However, one should also take into account other dimensions of interest, such as the geographical coverage. If the interest is in analyses of local or regional entities special considerations need to be made for the sampling procedure and the choice of reporting units. For the sampling procedure, for example, the question is whether the local dimension is important enough to be introduced as strata, in which case location would be 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.

2.24 To calculate statistics at subnational/regional level, information about the location of a firm should be known. This is often not very problematic because the contact detail of a firm is known anyway. The problem is what to do when firms are located at more than one location. This should be considered at the beginning of the survey implementation in order to decide how to proceed.

3. Sampling and reporting units

2.25 In planning BTS, it is useful to distinguish the enterprise and the establishment. An *enterprise* is an institutional unit in its capacity as a producer of goods and services (paragraph 5.1, 2008 SNA). An enterprise may be engaged in one or more economic activities at one or more locations. If enterprises are grouped together on the basis of their principal activities, at

least some of the resulting groupings are likely to be very heterogeneous with respect to the type of production processes carried out and also the goods and services produced. Thus, for analyses of production in which the technology of production plays an important role, it is necessary to work with groups of producers that are engaged in essentially the same kind of production. Thus a further partition into smaller and more homogeneous units of production is required.

2.26 An *establishment* is an enterprise, or part of an enterprise, that is situated in a unique location and in which only one productive activity is carried out or in which the principal productive activity accounts for most of the value added (paragraph 5.2, 2008 SNA). A *kind-of-activity* (KAU) unit is an enterprise, or a part of an enterprise, that engages in only one kind of productive activity or in which the principal productive activity accounts for most of the value added. A *local unit* is an enterprise, or a part of an enterprise, that engages in productive activity at or from one location. Table 2.2 shows schematically how the different units relate to each other in the case of an enterprise which produces three products and operates in three different locations.

Table 2.2: Breakdown of an enterprise

Enterprise	*Location 1*	*Location 2*	*Location 3*	
Product 1	*Establishment*	*Establishment*	*Establishment*	*KAU 1*
Product 2	*Establishment*	*Establishment*	*Establishment*	*KAU 2*
Product 3	*Establishment*	*Establishment*	*Establishment*	*KAU 3*
	Local unit 1	*Local unit 2*	*Local unit 3*	

2.27 A *statistical unit* is an entity about which information is sought and for which statistics are ultimately compiled, e.g. an enterprise, usually contained in the survey frame and used for sample selection. For operational purposes, a distinction is made between reporting and collection units (UN, 2007). The *reporting unit* is the unit about which data are reported.

Reporting units are those entities for which information is collected by means of questionnaires or interviews. The reporting unit may or may not be the establishment. Reporting units will, in most cases, coincide with the units for which statistics are compiled, as in the case of single-establishment enterprises where the enterprise and the establishment are identical. In the case of multi-establishment enterprises, the enterprise may make a separate return for each establishment, or each establishment may make a return for itself. A *collection unit* is the unit from which data are obtained and by which questionnaire survey forms are completed and provides answers about the reporting unit. In fact, it is more of a contact address than a unit. Sometimes the questionnaire is filled in by a central administrative office or an accountancy firm which provides this service to its client. Such information-providing entities constitute collection units. (UN, 2009).

2.28 The choice of the reporting unit can influence the survey results. If the reporting unit is the establishment, the aggregate by economic activity will be homogeneous in terms of economic activity covered and the aggregate for different regions will correctly reflect the activities in the region. If the reporting unit is the local unit, the regional aggregates will correctly reflect the activities in the region, but the aggregate by economic activities will likely cover different types of production activities, possibly classified for example in different Sections of ISIC Rev. 4. Finally, if the reporting unit is chosen to be the kind-of-activity unit, the aggregates by economic activity correctly reflect the economic activity, but the geographical representation is not guaranteed.

2.29 Ideally, the reporting unit for business tendency surveys should be the establishment; however, in practice, it is often difficult to obtain information at the establishment level: multi-establishment enterprises often cannot provide reliable information at the establishment level. Since BTS mainly focus on the information by different economic activities (for example, manufacturing, services, construction, etc.), the KAU is an acceptable alternative choice for the reporting unit as it allows to have reliable and meaningful aggregates by economic activity. Generally the questionnaire is sent to the manager of the enterprise who can either complete it for all the reporting units or send it to the reporting units for completion.

2.30 If location is an important element of the survey and the KAU is used as the reporting unit, methods can be used to identify the location of the units and “distribute” the responses of

KAUs to various locations according to size information for establishments, e.g. number of local employees.

4. Basic principles of questions

2.31 BTS have been conducted in various countries for many years. Over time common elements have emerged on the survey questions and the response scales. Details about the questions are discussed in Chapter 3 and in Annex 2. The selection of questions is based on the following guiding principles:

- The questions should measure developments in the business activity at an early stage (in order to timely capture business cycle developments);
- The questions should reflect topics that are sensitive to changes in the economic environment;
- The questions should mainly focus on assessments, expectations and plans even though some quantitative information may be asked.

2.32 The number of questions should also be kept to a minimum in order to reduce response burden. For these reasons, it is important to focus on a few key questions. This is important especially if the survey is to be carried out on a monthly basis. An approach that has been adopted by a number of countries is to use a very short questionnaire each month - consisting of about ten questions at most - and to add a few additional questions every quarter or half-year.

2.33 In addition to questions relevant to the business cycle, it may be useful to include also questions on bottlenecks/constraints in the questionnaire. This type of question asks about factors hindering the business activity and they are often asked less frequently than the usual questions. Questions of this type could be interesting for policy recommendations and for measuring reform progress especially in developing and emerging economies.

2.34 Another aspect which is important to achieve cooperation from respondents is to keep questions relevant from the perspective of the managers of enterprises and not only from a business cycle theory perspective. This is important since BTS are in most cases voluntary panel surveys where respondents are asked on a regular basis. Respondents are more likely to answer regularly when the survey questions (and their results) are relevant for them. In addition, using

appropriate language in the formulation of the questions - that is targeted to the respondents - may foster collaboration by respondents and ensure good quality of their answers. For example, if the targeted respondents are in the manufacturing industry, then the term *shipment* is more familiar than the term *sales* which would be more appropriate for respondents in retail trade.

2.35 With regard to the response scales, most of the responses in BTS are measured on pre-defined qualitative scales. Examples of pre-defined answer categories are: increase; remain unchanged; decrease or improve; deteriorate (for further details see Chapter 3).

2.36 Overall, the experience with BTS shows that they give timely and valuable information about economic activity. This information is demanded by various user groups both in countries with an advanced system of national accounts and economic statistics and in countries with a less advanced system.

5. Survey frequency

2.37 Behind the decision about survey frequency is a trade-off between the workload of the respondents and the institution conducting the survey and the need for timely signals about economic development. To monitor economic movements, indicators should have a monthly or, at most, a quarterly frequency. Biannual or annual data are more relevant for structural analyses than for monitoring the business cycle. For some policies, the signals provided by annual data often come too late: policy measures, like stimulus packages, need time for implementation, and therefore they need timely signals. From a user perspective a monthly survey is optimal and is the recommended frequency for conducting business tendency surveys. When this is not feasible, it may be a sensible way to start a business tendency survey programme on a quarterly basis with the goal of moving to monthly frequency as soon as possible.

C. The scope of Consumer Tendency Surveys

2.38 As for BTS, the first step in the construction of CTS is to define its scope in terms of target population, data needs, geographical coverage, variables to be inquired, type of the questions and frequency of the survey (see Table 2.1).

1. Target population, sampling and reporting unit

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

2.40 The sampling unit is usually an individual selected directly or within a given household (depending on the available frame list); the reporting unit is generally coincident with the sampling unit.

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

2. Principle of the questions

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

2.43 Consumers survey questionnaires usually contain a larger number of qualitative questions (15 questions – 12 asked monthly and 3 quarterly - in the Joint Harmonised EU Programme of Business and Consumer Surveys User Guide) generally allowing five possible replies arranged on a (generally 5-level) Likert scale ranging from the extremely positive to the extremely negative. The qualitative questions are preceded by a section aimed at gathering structural information on the respondent and his/her household. Structural information usually includes demographic characteristics of the respondent and his/her household, together with socio-economic characteristics of the respondent.

2.44 The qualitative questions usually ask for an assessment of the economic situation of the household and that of the country, saving and buying intentions, with particular emphasis on durable goods. Questions are usually referred either to the situation of the last 12 months or to that of 12 months ahead. Chapter 3 provides more detail on questionnaire development for CTS.

3. Survey mode and frequency

2.45 CTS are usually performed by telephone. Fixed lines are normally used; more recently, in some cases, mobile phones have been used too. The use of the internet is generally limited in the case of CTS: the use of the internet mode may bias survey results in cases where internet usage is not fully homogeneous across the population in a country. Moreover, internet surveys imply a stronger commitment from the respondents to reply. A good response rate is often more difficult to achieve for consumers than for firms, since the relevance of survey information is less immediate for the consumers than for firms. When telephone or internet cannot be used, CTS may be performed by face-to-face interview or by paper questionnaire (for further details about data collection modes see Chapter 6).

2.46 It is recommended that CTS be carried out on a monthly basis. Some of the questions (more specifically, those on the intention to buy durables) can be administered on a quarterly basis.

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 tendency surveys are generally released soon after the end of the reference month, making it easier for policymakers and analysts to assess 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 standardized at an international level. It is therefore relevant to define a frame of internationally agreed sets of questions used as a reference for countries. The Joint Harmonised EU Programme is a well-established example of such a framework (EC DG ECFIN, 2014) and it provides the basis for the OECD programme on business tendency survey as well for non-OECD countries. Some adaptations to this framework may also be considered to reflect specific country's situation.

3.3 This chapter addresses the question of how to design business and consumer tendency questionnaires in a standardized way. Section B starts with general considerations on the main differences between business and consumer tendency surveys. It provides guidelines for the formulation of questions and the methods of measuring the responses and a framework for the standardization of questions in their time dimension, scale and period covered across kinds of

economic activities. It also discusses the pre-testing phase of the questionnaire. Section C presents a set of core variables for different frequencies (monthly, quarterly and annual) for both business and consumer surveys. The standard questionnaires for BTS and CTS are presented in Annex 2.

B. General considerations

1. Business and consumer tendency surveys: main differences

3.4 While BTS and CTS have many elements in common (such as the frequency, timeliness, simplicity of the questionnaire), they have some main difference in terms of theirs scope and the type of information collected. The main differences are described below.

3.5 BTS are designed to capture information about the activity of an economic unit, following the entire business cycle from orders, production, and stocks to sales. The accuracy in terms of measuring the production process is rooted in the unit itself. Given that the targeted respondents are usually business managers, the answers are expected to be reliable to reflect a quantitative assessment rather than simply subjective perceptions of production. Questions refer to a relatively short time period (e.g. evaluating the current situation, assessing the recent past and near future developments): business managers have access to the information required for assessing the most recent past and provide an accurate estimation for the near future, which adds soundness to their responses.

3.6 CTS consist of questions dealing with the individual financial situation of households as well as their perception of 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. This information is obtained through a set of "micro-questions" focusing on households' financial situation, past, current and future and on saving intentions and capacity,

and questions asking if consumers are likely to make durable consumption expenditures in the present or in the future (consumers are asked, for example, to 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 quarterly survey, consumers are asked about the likelihood of buying a car or buying/building a home over the next 12 months). 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. Questions should be formulated clearly to avoid confusion by providing clear distinction between questions aimed at household level and those aimed at country level.

2. Formulation of questions

3.7 Reducing the response burden of a survey often leads to an increased response rate. Therefore, questionnaires should be designed to be self-contained in the instructions, straightforward in compilation and easily accessible. Their content should also be relevant to the respondents - in this case, business managers or 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 made clear to the respondents that the confidentiality of the information supplied is preserved and 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 for the past of future assessments, or type of questions, should be avoided in order to ensure international comparability. 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. As in the case of the Joint Harmonised EU Programme of Business and Consumer Surveys, a very short questionnaire should be used for the regular monthly BTS, while adding additional questions every quarter or half-year to monitor other aspects of the economy. The quarterly questions should always be the same and the main reason not to include them on a monthly basis is to alleviate the response burden on participants.

3.9 An important aspect to consider when designing a questionnaire is related to the wording of questions in national languages. Special care should be taken when translating the questions into national languages to adhere to the meaning of the original questions in English. In addition, as is better described hereinafter, all questions should be tested with a variety of respondents to determine whether they will lead to accurate responses.

3.10 Another important element for data comparability is the *timing* of conducting the survey. In particular, it is recommended that the fieldwork for monthly surveys be performed in the first two to three weeks of each month for monthly surveys and in the first two to three weeks of the first month of each quarter (January, April, July and October) for quarterly surveys. Data should be published no later than the end of the month following the reference month. Because of the importance of timely results of ETS for analytical purposes, countries are encouraged to make efforts to publish the results even earlier than the end of the following month.

3. Choosing topics and economic activities

3.11 The choice of information collected in business and consumer tendency survey 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 an incentive for their continuous participation in the survey.

3.12 In BTS, different types of information may be of interest to collect; these include the following:

(a) Expectations of future sales/ production/ business activity or situation[1], possibly also order books/ demand (turnover);

(b) Past sales/ production/ business activity or situation;

(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 surveys generally cover questions for each of these seven categories, but not necessarily each month, and not for all economic activities. Generally speaking, the choice of questions may be specific to the country situation. For example, in countries where companies do not have order books, questions related to them become irrelevant; for export-oriented countries, it may be useful to take into account factors limiting export orders which may affect production. In emerging industrial and developing economies, interruption in the supply chain of raw materials, fuel, electricity or other energy inputs is an important issue that has an impact on the production cycle. For these countries, factors limiting production should include the shortage of raw materials and energy (fuel and electricity). Other possible variables to be included in the

[1] Depending on the type of economic activities to which they belong, firms may be asked about their sales, their production, their business activity or their business situation (for further details see the questionnaires in the Annex 2). All these terms refer to output.

questionnaire are: the trend in the cost of labour, and, among the factors limiting production, the shortage of technology equipment, financial constraints, state legal policies, etc.

3.14 BTS are usually conducted for the following economic activities: manufacturing, retail trade, construction, and services (see Annex 1 for the correspondence with ISIC Rev. 4). Of course, different divisions or groups could be implemented depending on the budget allocated to tendency surveys and on the structure of the economy. A specific survey for agriculture could be considered, for instance, depending on the contribution of this kind of economic activities to GDP. In developed countries, it could also be useful to consider further breakdown of the economic activities. For example, service activities could be further broken down into more homogeneous divisions such as transport, business services and recreation services, since they are quite different from each other.

3.15 In CTS, two types of questions are generally considered:

- Micro-questions, and
- Macro-questions.

Micro-questions refer to 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. Macro-questions refer to questions regarding the general economic situation in the country, the evolution of unemployment and consumer prices, in the past and in the future. As for BTS, not all questions are necessarily asked in CTS on a monthly basis.

4. Measurement scale

3.16 The information collected in ETS is mainly qualitative with replies framed in a *Likert* scale type of measure (from three to five options). Multiple choice questions are used for most variables in almost all ETS. Generally, the number of choice questions depends on the variable

for which an opinion is asked and it is linked to the need to measure the intensity of the actual and predicted change.

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

- Above normal / normal for the season / below normal
- More than sufficient / sufficient / not sufficient
- Increase / remain unchanged / decrease
- Improve / remain unchanged / deteriorate

3.18 The first two sets of options are only used in questions assessing the present, as the evaluation scale requires no comparison to the past or the future. Moreover, the first set of options is usually used for questions asking the entrepreneurs to judge the actual level of certain endogenous variables in relation to the level that is usually called normal for the season. It is worthwhile mentioning that the notion of normality is not specified by the survey and it is left to the respondent to give his/ her own subjective meaning.

3.19 There may be some exceptions in BTS where questions are not formulated around a three-option scale. In a few cases four or, at most five, choices may be offered. These questions call for a judgement on recent developments, an assessment of the current situation, or expectations for the near future. In addition, there may be quantitative questions, such as on capacity utilisation, which refer to a percentage of full capacity utilisation, and on the number of months of production assured by the orders at hand, which, of course, refers to a number of months. There may also be questions asking about which factors are limiting production, activity or investment based on a list of factors to choose from. Finally, on a biannual basis, there are questions on the structure of investment.

3.20 In CTS, questions are generally formulated around a three[2]/ four/ five option scale. In the Joint EU Harmonised Programme, typical possible answers are around four or five options. When questions are formulated around a four option scale, the neutral answer is generally not included in the available options. Examples of possible reply options are listed below:

- Very likely / fairly likely / not likely / not at all likely
- Yes, definitely / possibly / probably not / no
- A lot better / a little better / the same / a little worse / a lot worse
- Much more / a little more / about the same / a little less / much less
- Increase more sharply / increase slightly / remain the same / fall slightly / fall sharply

3.21 Note that in the questionnaires used for CTS, there is a reply option for "do not know".

3.22 As the information collected in ETS is mainly qualitative, one needs to find a way to interpret it. Percentages obtained from the multiple-choice questions may be estimated according to the methods described in Chapter 5; these values are usually converted into a single number by using net balances or diffusion indices. Computational details of net balances and diffusion indices, as well as alternative methods of quantification of the replies, are provided in Chapter 8.

5. Specifying the time dimension of the questions

3.23 In order to achieve comparability over time, across countries and across questions, it is desirable to use the same time dimension as well as the same measurement scale in the questionnaire. The time dimension in ETS relates to the recent past, the current situation or the near future.

---

[2] For some questions, a three option scale is used in the Surveys of Consumers carried out by the Survey Research Center at the University of Michigan. For further details about the questionnaire: http://www.sca.isr.umich.edu/fetchdoc.php?docid=24776.

3.24 In monthly and quarterly BTS, a time span of three months is considered a benchmark in order to assess past and future evolutions as well as the present situation. The time span of three months is suitable to capture the direction of change for a specific activity. Moreover, it is consistent with managers' capacity to think in quarter-to-quarter terms taking into account past decisions. However, when questions are added to the standard questionnaire (to address specific topics of interest for a country), the time-horizons may vary. An example is the question on business expectations by the IFO Institute for Economic Research (Germany) in Germany which uses a time horizon of the next six months. Answers to this question are then used to compute the business climate in Germany[3]. The time span may also be different for semi-annual surveys. For example, in the Joint EU Harmonised questionnaire for BTS, the semi-annual investment survey asks manufacturing firms about the evolution of their investments in the previous year and about their expectations for the year to come.

3.25 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 account 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 a larger time-span is to reduce the volatility of the responses.

6. Seasonality

3.26 The seasonal component is that part of the variations in a time series representing intra-year fluctuations that are more or less stable year after year with respect to timing, direction and

[3] 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

magnitude[4]. 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 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.

3.27 Firm managers are generally aware of the seasonal patterns affecting their business production, sales, stock levels, etc. It is therefore important to inform the respondents on whether or not they should take seasonality into account in their answers. This information can be given at the beginning of the questionnaire, as a general instruction, 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 tendency survey data shows that this tends to reduce seasonality but does not eliminate it entirely. It is recommended that respondents be asked to exclude seasonal variations and that all data, after having been converted into net balances or diffusion indices, should be systematically tested for seasonality and seasonally adjusted when needed[5].

---

4 OECD Glossary of Statistical Terms.

[5] Seasonal adjustment methods suitable for ETS series, including computational details on net balances and diffusion indices, are treated in chapter 8.

7. Characteristic information of respondents

3.28 Business and consumer tendency surveys are structured in order to collect qualitative information regarding the company and the consumers' perceptions, as well as to provide a set of information on the characteristics of the company and the household participating to the survey. As qualitative information regarding companies and households has been described in the previous paragraphs, this section lists the required information on both the participating company and the household.

3.29 BTS should enquire on the following items of a company:

(a) Kind of economic activity/ Principal economic activity

(b) Company name

(c) Company address

(d) Total number of persons employed

(e) Name and position of respondent

(f) Contact information (telephone number, facsimile number, email address)

(g) Total revenue (for this item, relevant categories for responses should be provided)

3.30 Similarly, CTS should collect the following information per household surveyed:

(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[6]

a. in Employment

b. in Unemployment

c. Outside the labour force

(vii) Form of work[7]

a. mainly in own-use production work

b. mainly in employment

c. mainly in unpaid trainee work

d. mainly in volunteer work

e. mainly in other forms of work

---

[6] The definitions are based on the Labour Force status of the *19th International Conference of Labour Statisticians*. See http://www.ilo.org/wcmsp5/groups/public/---dgreports/---stat/documents/normativeinstrument/wcms_230304.pdf.

[7] See footnote 6.

f. exclusively in non-productive activities

(e) Household characteristics

(i) Household size (for this item, relevant categories for responses should be provided)

(ii) Number of employed persons

a. Full time

b. Part time

(iii) Number of overseas workers[8]

(iv) Household income (for this item, relevant categories for responses should be provided)

(f) The status of housing ownership

(i) Owned

(ii) Rented

(iii) Free

(v) Other

8. Pre-testing of the questionnaire

3.31 Pre-testing of the questionnaire is a very important part of a questionnaire design. The procedure consists of a systematic check of the survey questions, questionnaires and other survey procedures prior to the full scale implementation of the survey in order to identify possible problems with its execution and solutions before the official release. Rather than reporting results, pre-testing serves the purpose of checking for possible problems in survey implementation. For example it helps assess whether the questions are relevant and easily

[8] This may apply to developing countries only.

understood by the respondents, if questions order is appropriate, etc. Pre-testing is also useful to determine the duration of interviews and is a good tool to assess respondents' willingness to participate to the survey.

3.32 Usually if some issues occur in the pre-testing phase, it is likely that similar problems will arise during the full-scale administration of the survey. 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 to determine whether there are changes needed to improve the survey instrument.

3.33 Pre-testing techniques are divided into two major categories: *pre-field* and *field* techniques. Pre-field techniques are generally used during the preliminary stages of questionnaire development and include: respondents' focus group and cognitive laboratory interview. Field techniques test questionnaires under operational conditions. These techniques include behaviour coding of interviewer/ respondent interactions, interviewer/ respondent debriefings, split-sample tests, and analysis of item non-response rates and response distributions. Both pre-field and field testing should be done when time and financial resources permit. However, in some situations it is not feasible to use all methods. In these cases, a combination of methods can provide good results for identifying and resolving problems. The following should be considered when conducting the pre-testing:

- Detailed notes on encountered problems should be included and possible solutions should be identified.

- The pre-test of the questionnaire should be carried out also with specialists in questions design, who may be able to identify potential issues otherwise difficult to see in a pre-test with respondents only.
- The whole variety of respondent types 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.
- Specialists in the subject matter should be present at pre-testing. They should be able to determine if respondents understood the question properly. For electronic questionnaires, it is important to test the electronic application with the respondent in order to determine if functionality, navigation, etc. are user friendly.

3.34 Based on the results of the pre-testing, potential modifications that can be made to the survey include: revise the structure of the questionnaire, such as the order or sequence of questions, filters, etc.; add more instructions to the questionnaire and interviewer's manual; delete redundant questions / add new ones; and reword or rephrase questions. After resolving the problems and issues encountered during the pre-testing, the next step is to finalize the survey questionnaire ready for implementation in the full-scale survey.

C. Standard questionnaires

3.35 This section presents an overview of the type of questions used for business and consumer tendency survey. As mentioned in the introduction of the Handbook, the recommendations are based on the experience of countries and on the work of both DG ECFIN (see EC DG ECFIN, 2014) and the OECD (see OECD, 2003) on tendency surveys. The following example draws heavily on the Joint Harmonised EU Programme of Business and Consumer Surveys which is considered as a reference for developed countries even outside of

the EU. Annex 2 presents a detailed list of questions for BTS and CTS. As a general rule, it is recommended to use in the tendency survey the minimum set of questions (monthly and quarterly) as described in the tables of this section and Annex 2, to facilitate the comparison among countries. Countries can introduce additional questions to those presented in this handbook based on specific topics of interest.

3.36 For all questions, respondents should be requested to exclude purely seasonal fluctuations from their responses. After having been converted into net balances or diffusion indices, answers should also be seasonally adjusted when needed (see also section B.6). If a question has been introduced too recently for traditional seasonal adjustment methods to be implemented, one can temporarily ask respondents for additional comparisons with the year before.

1. Monthly business tendency surveys

3.37 Table 3.1 shows the list of topics which are used in monthly BTS together with the relevant time dimension: questions may refer to the past or to the future, in which case three months should be considered as a benchmark, or the present.

3.38 Questions referring to the evolution of production, business activity or sales, depending on the kind of economic activities, over the last 3 months, allow for the now-casting of economic activity (e.g. GDP) before national accounts figures are released. Those referring to its evolution over the next 3 months allow forecasting economic activity.

3.39 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 3 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.

Table 3.1: Topics of questions for monthly business tendency surveys

Economic activity	Topic of the questions	PAST (last 3 months)	CURRENT	FUTURE (next 3 months)
MANUFACTURING	Evolution of production	**x**		**x**
	Evolution of employment	x		**x**
	Level of order books		**x**	
	Level of export order books		**x**	
	Stock of finished goods		**x**	
	Evolution of selling prices			**x**
CONSTRUCTION	Evolution of building activity	**x**		x
	Evolution of employment	x		**x**
	Level of order books		**x**	
	Factors limiting production		**x**	
	Evolution of selling prices			**x**
RETAIL TRADE	Evolution of business activity(sales)	**x**		**x**
	Evolution of employment	x		**x**
	Evolution of orders placed with suppliers			**x**
	Stock of finished goods		**x**	
	Evolution of selling prices			**x**
SERVICES	Evolution of business situation	**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	x		x
	Evolution of employment	x		x
	Factors limiting production		x	
	Stock of finished goods		x	
	Evolution of selling prices			x
FINANCIAL SERVICES	Evolution of business situation	**x**		x
	Evolution of employment	**x**		**x**
	Evolution of demand (turnover)	**x**		**x**

2. Quarterly business tendency survey

3.40 Quarterly questions are added for most kind of economic activities to supplement the monthly topics. The reason to ask these questions at a different frequency is to reduce the response burden of monthly questionnaires. Also, some variables typically do not need to be asked every month. Questions, for example, on factors limiting production and on the assessment of production capacities are usually asked on a quarterly basis because they are not prone to rapid changes. As for monthly questionnaires, three months is considered a benchmark time span for all questions referring to the past or to the future. Table 3.2 provides a list for the different economic activities.

Table 3.2: Topics of questions for quarterly business tendency surveys

Economic activity	Topic of the questions	PAST (last 3 months)	CURRENT	FUTURE (next 3 months)
MANUFACTURING	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	
SERVICES	Factors limiting activity	**X**	
	Increase of volume of activity with present resources	**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**

3. Semi-annual business tendency surveys

3.41 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. Questions on investments can be added to the monthly BTS on a semi-annual basis, or covered in a separate survey. The latter is the approach taken in the Joint Harmonised EU Programme, where a survey on investments is conducted twice a year separately from the monthly survey. The EU survey covers both quantitative questions on the relative evolution of investment in the last year and in the next year (% change[9]) as well as qualitative questions on the structure of investment and on factors influencing it (e.g. demand, financial resources or expected profits, technical factors, other factors).

3.42 In order to reduce the response burden, this handbook recommends including quantitative questions regarding the evolution of investment in the monthly BTS on a semi-annual basis. In the 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

[9] Some EU countries (e.g. Germany, Denmark, Italy etc.) do not ask about percentage changes. They ask about actual amount invested. For each firm, percentage changes are calculated in the processing data step.

between year *t* and *(t-1)*. In the second survey, conducted six months later, firms would be asked for the change in investment between years *t* and *(t-1)* and expectations for change in investments between years *(t+1)* and *t*. Questions on investments should be considered, at least, for Manufacturing and other economic activities contributing significantly to the evolution of investment aggregates in the country.

4. Monthly and quarterly consumer tendency surveys

3.43 Questions in CTS can also be presented according to the time dimension they refer to, namely evolution in recent past, assessment of current situation, and expectations for the near future. The time span used in CTS is usually larger than that of BTS. It is in fact common for consumers to look at the 12 months back and forward (see also section C.2). Table 3.3 and Table 3.4 present the core topics of questions for monthly and quarterly consumer tendency surveys.

Table 3.3: Topics of questions for monthly consumer tendency surveys

Topic of the questions	Past (last 12 months)	Current	Future (next 12 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 3.4: Topics of questions for quarterly consumer tendency surveys

Topic of the questions	Past (last 12 months)	Current	Future (next 12 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**

3.44 Additional questions could be considered. For instance, the US Surveys of Consumers carried out by the Institute for Social Research at the University of Michigan, also includes questions about the subjective probability assigned by consumers to the realization of specific events (income growth being higher or lower than inflation, falling into unemployment, adequacy of the retirement scheme) and questions on interest rate expectations and price expectations for specific goods (gasoline and house prices).

Chapter 4: Survey frame and sample design

A. Introduction

4.1 While the previous chapters deal with the general scope and use of ETS and their desired output, Chapters 4 and 5 discuss the inputs that are needed to produce ETS data, and how these inputs must be processed to arrive at the final output. More specifically, this chapter focuses on important steps in survey process: identifying the survey frame and the sample design, that is how to select a subset of units (i.e. the sample) from the frame based on which we can draw estimates for the target population of the information we wish to collect (i.e. percentages of reply options for questions composing confidence indicator).

4.2 Census surveys covering all business/consumers are not a feasible option for conducting tendency surveys as they represent a time consuming and resource intensive exercise to collect data on a monthly and quarterly basis. Sample surveys are used instead to collect data from a smaller sample of units, which requires a sample design - the main topic of this chapter.

4.3 There is no unique or optimal sample design which is applicable in all circumstances. The sample design depends on the specific situation, including the availability and type of the survey frame, the objectives of the survey, the structure of the economy and the population, the financial and human resources available, etc. This chapter aims at providing an overview of the methods available for economic tendency surveys. Countries should decide on the appropriate method based on their country-specific situation.

4.4 Section B provides a description of the survey frame and describes how to identify the appropriate frame for conducting tendency surveys. Section C presents specific characteristics of the sample design that should be taken into consideration when designing the sample survey.

B. Survey frame

4.5 In setting up a survey the first thing that needs to be addressed is to clearly identify the population from which to collect information. This is generally defined as the target universe and may comprise all or only a subset of interest of the overall population. In the case of business surveys, for example, the target universe consists of all the companies/firms carrying out some specific economic activities, such as, manufacturing, services, etc. For consumer surveys, the target population consists of individuals belonging only to some age groups (e.g. full age persons).

4.6 The survey frame is a tool that is used to identify and have access to units in the survey population. It is used to select the sample and it is an important element in the design of a survey as the chosen frame will impact the scope of the survey population. For example, if the survey frame consists of a list of telephone numbers, households without a telephone would be excluded from the sample. The sampling unit is an element of the survey frame, thus subject to being selected in the sample. Of course, the choice of the target population, survey population, survey frame and sampling unit are defined in conjunctions with one another.

4.7 Choosing the right survey frame impacts the results of the survey; it is thus important to have a good survey frame for which the survey population is as close as possible to the target population. A good survey frame should be reliable, complete and up-to-date and it should not contain duplicated or non-relevant elements and, as far as possible, has to include all units of the investigation.

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

4.9 Mention should be made of *multiple-frame sampling,* which refers to surveys in which two or more survey frames are used and independent samples are respectively taken from each of the frames. Inferences about the target population are based on the combined sample data; this method is referred to as dual-frame sampling when the survey uses two frames. Multiple-frame sampling is used to achieve a desired level of precision with reduced cost and to have a better coverage of the target population. Even if a complete frame, such as a household address list, is available, it is often more cost-effective to take a sample of reduced size from the complete frame and supplement the sample with additional data taken from other frames, such as telephone directories or institutional lists that might be incomplete but less expensive to sample from (Hartley, 1974, Lohr and Rao, 2006; Skinner et al., 1994).

1. Business tendency surveys

4.10 Once the sector of interest has been identified, the target population of a business tendency survey consists of all economic units engaged in the relevant kinds of economic activity, regardless of their type of ownership, legal form, size, and institutional form. In addition, the target population should cover all enterprises active at any time during the period covered in that survey (see also Chapter 7).

4.11 As the selection of the survey frame is affected by the decision on which sampling units to use, the different types of economic units are described next (see definitions in Chapter 2).

4.12 An *enterprise* often engages in productive activities at more than one location, and, for some purposes, it may be useful to partition it accordingly. Thus, a *local unit* is an enterprise, or a part of an enterprise (for example, a workshop, factory, warehouse, office, mine or depot), which engages in productive activities at or from one location (e.g. all establishments at the same location together constitute a local unit);

4.13 For *kind-of-activity* units, in contrast to establishments, there are no restrictions on the geographic area in which the activity is carried out (e.g. all establishments carrying out the same industrial activity together constitute the kind-of-activity unit). Many companies, especially the smaller ones, are located in just one location and their whole activity is concentrated in one particular kind of activity. For these the establishment, local unit, kind of activity unit and enterprise are the same.

4.14 The choice of the sampling unit has the following implications:

- 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 sampling 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. A trade-off between the available survey frame and the output specification is often necessary. (See also OECD, 2003).

4.15 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. As mentioned in Chapter 2, since BTS mainly focus on the information by different economic activities (for example, manufacturing, services, construction, etc.), the KAU is an acceptable alternative choice for the reporting unit as it allows for reliable and meaningful aggregates by economic activity. Generally the questionnaire is sent to the manager of the enterprise, who can either complete it for all the reporting units or send it to the reporting units for completion.

4.16 The survey frame is usually based on 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 in terms of 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. Detailed guidance in designing and maintaining business registers is available in Eurostat Manual on business registers (Eurostat, 2003) and the upcoming UNECE Guidelines on Statistical Business registers (UNECE, 2015). Here we assume that there is a reasonably comprehensive business register and that, at minimum, it contains information on the addresses of enterprises, their main kinds of activity and their approximate size.

4.17 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 comprehensive business registers of the kind that national statistical agencies maintain for their regular enterprise surveys, if available.

2. Consumer tendency surveys

4.18 In CTS the target population usually represents only a subset of the entire population of a country. Cut-off on the basis of age is usually applied, which varies across countries (with a minimum age of 14-18 years, to investigate only full age individuals, and a maximum age in some cases is applied too). Cut-offs on geographical basis or working status can also be applied.

4.19 As confidence is indirectly monitored by asking questions regarding the present and expected situation of the consumer himself/herself and of his/her household, some uncertainties in defining the reference population may arise, particularly whether to consider the consumer or

the household to which he/she belongs or both as sampling unit. This question has consequences on the frame to use, on the kind of sample to select, and on the estimation of the results.

4.20 When the sampling unit is represented by the consumer - intended as a full age person contributing also in non-monetary terms to the household management, the sampling unit coincides with the statistical unit: the individual who is interviewed is directly selected from the frame. However, when the sampling unit is represented by the household, an individual within the household needs to be further selected once the sampling unit is selected.

4.21 In consumer surveys, survey frames can contain lists of individuals such as population registers or electoral rolls, but could also refer to individuals from a mobile phone register. On the other hand, lists can deal with dwellings/households, such as fixed phone registers. The choice of the list, apart from the subject of investigation, can be affected by other considerations, like the availability of such listings and to the selected data collection mode. If, for example, 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 through face-to-face interviews, a population register will be the most advisable survey frame.

4.22 Fixed telephone lists still represent the most used frame for consumer surveys. They are however affected by some drawbacks which affect the sampling design: the fixed phone directories refer to dwellings/households and not to individuals. Furthermore, in these lists increasing risks of under-coverage are present. These risks include the possible exclusion of part of the population who cannot afford the phone cost, or who chose not to be included in the telephone directories because of privacy reasons. In addition, with the increasing usage of mobile phones, another reason of under-coverage is due to the growing switching to mobile phones.

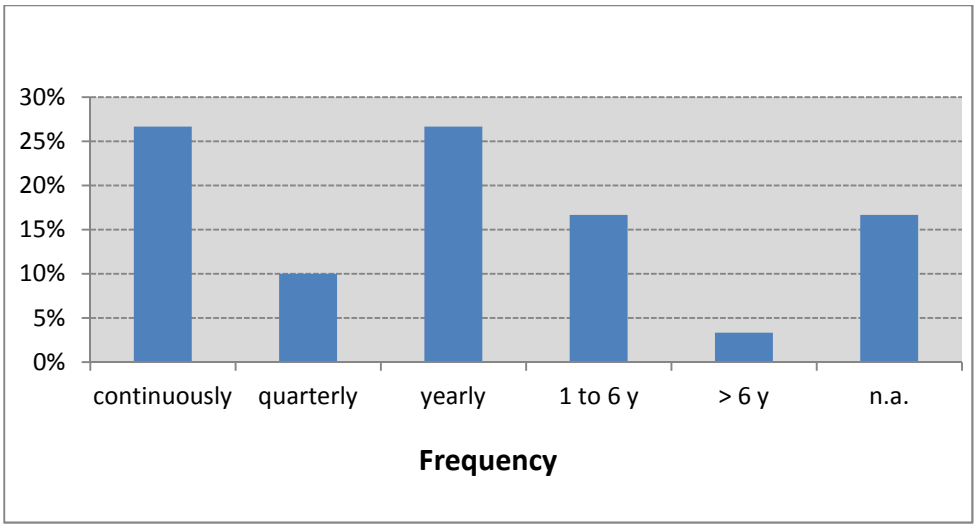
4.23 Mobile phones also have some limits, as pointed out by R. Curtin (2003, 2010). Generally, an interview conducted via mobile phone is more difficult, as respondents can be busy and not able to answer, also with risks for their safety (e.g. contacted while driving). Mobile phones are less used by 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 fixed to mobile - is increasing, it is not possible anymore to recognize the geographical location of the phone number, preventing possible stratification of the frame lists.

4.24 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.

4.25 Even though internet surveys are strongly increasing, lists based on internet connections presently show several drawbacks, most notably 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.

4.26 According to 2012 information, in OECD countries about 2/3 of institutes update consumers frame lists at least yearly (of which 27% do so continuously and this quota has even been 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 4.1).

Figure 4.1: Frame list update



Source: prepared by Istat based on EC Metadata and OECD _MEI

C. Sample design

4.27 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 accurate these estimates are. Only in the case of probabilistic samples (see below) can the accuracy of these estimates be determined.

4.28 The sample design comprises various features so as to increase efficiency and to accomplish the research aims. These features mainly consist of 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. This section provides a concise description of these aspects. 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 topic are referred to the literature.

1. Selecting units: probabilistic and deterministic samples

4.29 *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. This allows normality assumptions and consequently to calculate the accuracy of estimates (see chapter 5). The random criterion for selecting sample units is the principle on which the majority of ETS samples rely.

4.30 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 do not need to be calculated separately (see chapter 5).

4.31 In Section C.2 a concise but exhaustive discussion of the most applied samples designs based on probabilistic selection, best suitable for ETS, is presented.

4.32 For sake of completeness, a short introduction to *deterministic samples* is also 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 a specified accuracy. Furthermore, even a sample initially built as a probability sample sometimes ends up being a non-probability one, due to unintentional or unavoidable addition to the selected sampling design.

4.33 A list of deterministic samples used for consumer surveys is the following:

(a) The *purposive* or *judgmental sample* is a sample where 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;

(b) The *quota sample* represents the most used of the non-probabilistic methods, mainly applied in opinion polls and market research. Quotas are built by subdividing the population into homogeneous classes according to certain 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;

(c) The *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;

(d) The *Random Digit Dialing (RDD) techniques* refer to 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 University of Michigan is reported in Curtin (2013).

2. Sample design features

4.34 Many different sample designs exist, as they depend on the particular circumstances for which the design is the most efficient. Some often used sample designs based on probabilistic unit selection with their most important properties are described below.

4.35 *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, this is in practice automated, 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 is not 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 as benchmark to test their efficiency.

4.36 In *Stratified Random Sampling* two main conditions must be met: there must be a good degree of completeness of frame coverage, and the stratifying variables need to be strongly correlated with the variables of interest. In practice, the frame is divided first into several non-overlapping sub-frames (called strata) according to the chosen stratifying variables, and within each stratum a SRS design is used. Although conceptually simple, arriving at an efficient

Stratified Sampling design can be quite complex in practice: one needs to consider carefully not only 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 below).

4.37 Stratified Random Sampling can be used for various reasons:

(a) 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;

(b) 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 estimates to be made;

(c) 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;

(d) 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.

4.38 If the output specification requires a minimum level of precision for the estimators, the necessary stratum sample sizes can be determined using the methods discussed in chapter 5. Collectively, these sub-samples form the total sample. Often, however, the total sample size will be fixed, and the question is how to distribute the total sample size over the strata. This is known as the allocation problem. Different allocation methods exist in the literature; the two more applied allocation methods are the proportional allocation and the optimum allocation.

4.39 In *proportional allocation*, the size of the sample in each stratum is taken in proportion to the size of the stratum. If, for example, 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 (see also Chapter 5) all units have the same weight in the estimation. This case is referred to as self-weighting sample.

4.40 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.

4.41 When the variances within various strata 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 of 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 the Neyman allocation.

4.42 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; which simply means that the entire stratum has to be included in the sample.

4.43 Determining the optimum allocation requires that all stratum variances be 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 are 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.

4.44 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.

4.45 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 relationship 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"[1]).

4.46 Conceptually, a PPS design can be viewed as a sampling procedure which follows these steps: first cut 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 write the name of the unit on it. Then lay these strips of paper end-to-end on the floor in a random order so 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. Then measure out a distance equal to this random number from the left-most edge of the ribbon and mark it on the ribbon. Repeatedly make additional marks on the ribbon, each one a step length to the right of the previous mark, until the end of the ribbon is reached. 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.

4.47 This procedure only results in the correct number of units if no strips of paper are longer than the step length. A strip of paper longer than the step length would receive multiple marks. It is implicitly 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.

---

[1] Based on the experience of the OECD countries, a rule of thumb is that about 30 units are sufficient to obtain an acceptable level of precision for each of the strata for which data are to be published (for further details see OECD, 2003, paragraphs 65-66).

4.48 It is worth noting that PPS can be used when weights of selection are available for units chosen as sampling units.

4.49 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.

4.50 Other types of samples that can be used for ETS, in particular for consumer surveys, are the following:

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

(b) *Cluster sampling* is used when relatively homogeneous groupings (clusters) can be identified in the population where each cluster is a small-scale representation of the total population. A random sampling technique is used to select the clusters and all the units within the cluster are included in the sample (single-stage cluster). In a two-stage cluster sampling (a simple case of multi-stage sampling), instead, a random sampling technique is applied to select units from each of the selected clusters. Clustering is adopted in survey sampling for practical reasons. Mainly, if the sample needs to be reduced into a manageable workload for interviewers when, for example, using the face-to-face data collecting mode. This sampling technique, while allowing for saving travelling time and consequently reductions in costs, implies larger sampling error than in un-clustered samples since only a part of the population is monitored.

4.51 Less widespread sample 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 to use for BTS. Once the first sampling unit is randomly selected, instructions are given to the interviewer to follow a random route to interview individuals, for example, take the first road on the right, interview at second house on the left, continue down the road and interview at every tenth house on the right, etc.;

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

3. Panel sample

4.52 Generally, ETS are used to keep track of changes over time of variables of interest. Moreover the results from ETS provide early indicators for changes in the business cycle.

Therefore, it is advantageous to choose a survey strategy in which changes over time are measured accurately.

4.53 In this regard, *panel samples* provide a useful sampling method. A panel sampling is a method of selecting an initial sample (through any of the methods explained before) and surveying the same set of units over a period of time (e.g. every month and/or quarter). If all the units are maintained the sample is called *fixed panel*, if part of units are periodically substituted, it called *rotating panel*. Using a fixed panel rather than selecting a fresh sample each round of the survey increases the precision of estimates of changes over 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 widely used for business surveys, but less for consumer surveys.

4.54 In CTS, in fact, the panel sample may prevent asking consumers some sensitive questions such as those linked to income, or the household's economic situation. The surveyed consumer, in these cases, may feel controlled and will likely refuse to respond repeatedly to these kinds of questions.

4.55 In BTS, there are instead great practical advantages in using panel samples 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.

4.56 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 practical advantages 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.

4.57 The rotating panel design is also applied in the Surveys of Consumers of the University of Michigan. For the 500 units composing the sample about 60 per cent are new respondents, and 40 per cent are interviewed for a second time, after a time period of six months. This design, according to Curtin (2013), has several distinct advantages over non-panel samples for consumer surveys as well. It allows for better detecting of the underlying causes of the changes highlighted by the panel structure and it makes longitudinal analyses possible by repeated measurements. As the interviewed units are recurrent, the non-sampling influences remain relatively constant across waves. It is worth remembering, however, that even if the panel has been initially randomly selected, it cannot be considered a random sample in the subsequent waves.

4. BTS: specific features of sample design

4.58 An often used sample design for BTS is the stratified random sampling. One commonly stratifies according to the properties of the units in the survey frame. Each stratum is 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 for 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 economic activity;
- number of employees (or another measure of size, e.g. sales);
- geographical location.

4.59 When stratifying one also needs to decide the depth of stratification, namely: the degree of similarity of the units in the same strata; and how many different strata are allowed. For example, should just two categories for number of employees (e.g. “large” vs. “small” businesses) be distinguished, or should a more refined stratification be used?

4.60 The choice of the variables to use for the strata depends on a number of factors. Obviously the required auxiliary information must be available for all units in the survey frame. Also, the choice of the sampling unit will sometimes affect the choice of the stratification variables. For instance, the geographical location should not be used if the enterprise is used as the sampling unit, because an enterprise may consist of establishments in different locations.

4.61 The output specification should be taken into consideration when dealing with these questions. For example, if the output of different kinds of economic activity is required, it would make sense to stratify according to kind of economic activity. Moreover, it often does not make much sense to use a stratification scheme that is much more detailed than the output scheme.

4.62 The resulting number of strata should also be considered, especially when stratifying according to more than one variable at the same time. For example, if the stratification is done according to both kind of economic activity and geographical location and 30 kinds of economic activity and 40 regions are distinguished, the resulting number of strata is $30 \times 40 = 1200$ strata. Such a high number of strata may be prohibitive, given the maximum allowed sample size and the necessity for sampling at least a minimum number of units per stratum.

Box 4.1: Stratified sample in the Philippines

An example of stratified sample for BTS is the sample used in the Philippines for the quarterly BTS. The survey is conducted nationwide among all the regions in the country; the frame is a combination of two lists: Securities and Exchange Commission's (SEC's) top 7,000 corporations and Business World's top 1,000 corporations.

The SEC is the national government regulatory agency mandated to register corporations, collect fees from these corporations and prescribe reportorial requirements while Business World is the country's leading business newspaper. The rankings of the corporations or companies from both lists are based on gross revenues. The survey uses a stratified random sampling design and the sampling units are the companies extracted from the lists of SEC and Business World.

The combined list is subdivided into 192 strata classified by region (National Capital Region and Areas Outside National Capital Region for the other 15 regions of the country) and by 12 industry divisions, namely: (a) Group I: Industry - manufacturing; mining and quarrying; electricity, gas and water; agriculture, fishery and forestry; (b) Group II: Construction; (c) Group III: Services - financial intermediation; real estate, renting and business activities; hotels and restaurants; transport, storage and communications; community, social and personal services; (d) Group IV: Wholesale Trade and Retail Trade.

The industry classification of the corporations is based on the Philippine Standard Industrial Classification (PSIC). For the sample selection, a simple random sample of firms is drawn from each stratum for a total sample size of around 1,600 firms. For regions with a small number (less than 100) of top firms, a census survey is provided in all the strata. For the other regions, random samples were drawn from each stratum.

5. CTS: specific features of sample design

4.63 The consumer sampling design usually stems from a combination of different kinds of allocations techniques and selection of units; therefore the consumer samples usually present a complex sample design. The main features of sample design for consumer tendency surveys are presented in this section. For a more detailed discussion, the reader can refer, for example, to Cochran (1977) and Särndal et al. (2003).

4.64 In some countries consumer surveys rely on already existing social surveys by making use - even with some adjustment - of the same samples for cost effectiveness and also to take advantage of well consolidated experience.

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

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

4.67 Most countries conducting consumer tendency surveys have designed their sample to be representative in the sense that every adult member of the population has an equal chance of

being selected. In other words, the sample is selected according to an Equal Probability of Selection Method (EPSEM).

4.68 The EPSEM clustered (multi-staged) stratified random sample is one of the most widely used sampling techniques. This kind of sample design is adopted in survey sampling for practical reasons. When the sampling frame units cover two or more survey units (e.g. consumers within households), clustering is the only practical way of selecting a sample of the required units, or if the sample needs to be divided into manageable workloads for interviewers when using the face-to-face data collecting mode. While the multi-stage sampling allows for saving travelling time and consequently a reduction in costs, it implies a larger sampling error than un-clustered samples, as only a part of the population is monitored.

4.69 While the sample should be designed to be representative of the whole target population, some countries, such as China, Malaysia, and Mexico, conduct CTS based on a samples restricted to selected urban areas. This would not be a critical omission if the areas excluded from the sample do not fully participate in the economies of their countries (Curtin, 2007).

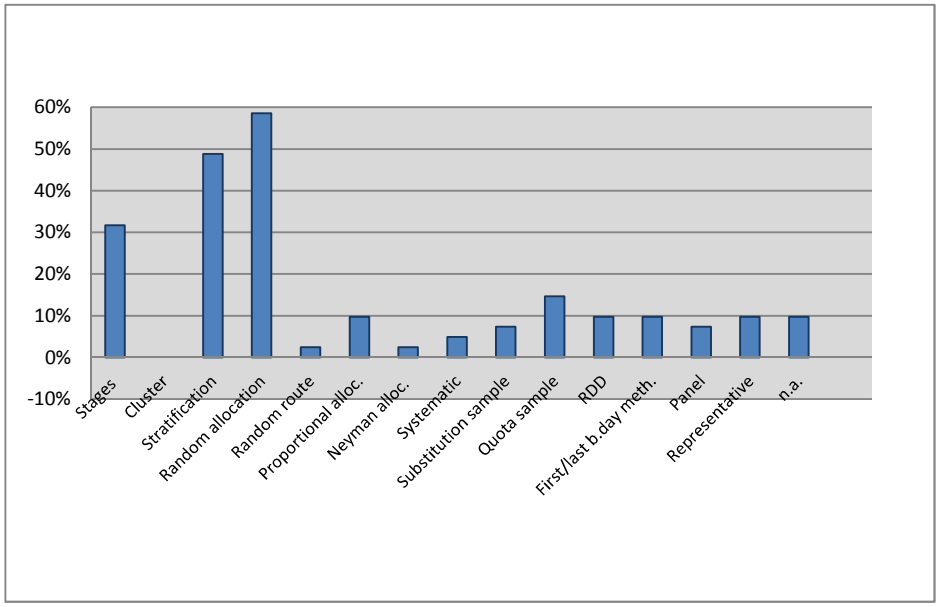
4.70 Figure 4.2 shows the main features of the consumer design samples available for the OECD countries, EU countries, and EU candidate countries. Various sample methods are used and the large majority of the samples are complex and built as multistage stratified random samples. The choice of the sampling methods used by countries reflects the specificities and capabilities of the different countries. In general, the sample designs of the countries in Figure 4.1 consist of least two of the characteristic reported in the figure.

4.71 The ISTAT CTS sample is an example of a complex sample design for consumer surveys. It is a stratified, two stage (telephone subscribers/consumers) random sample of exactly 2000 units, newly selected every month. The reference population is represented by the full age

(18+ years) population, stemming from the Census survey and updated yearly 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).

4.72 The frame is made up of the telephone subscribers list, structured by NUTS2 (regions), LAU2 and zip area codes; the primary sampling unit is the telephone subscriber selected, within each stratum, using a systematic random selection technique. The second stage sampling unit is the consumer. The consumer is selected from the adults belonging to the household which corresponds to the chosen telephone number. The selection criterion applied is the First Answer method, which is implemented 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 noting that the strata sample size exactly reflects the population proportion, but not that of the households. This occurrence affects the calculation of the inclusion probability (see Chapter 5).

Figure 4.2: Sampling design features of consumer surveys in different OECD, EU and EU candidate countries



Source: prepared by Istat based on EC Metadata and OECD _MEI

4.73 The sample design used in the Philippines for CTS is another example. It uses the 2003 master sample[2] of the Philippine Statistics Authority as a survey frame. The list of households in the master sample is considered a representative sample of households nationwide. The master sample was generated using a stratified multi-stage probability sampling scheme. The survey adopts the stratified multi-stage sampling design of the master sample in which the individual barangays[3] or combination of barangays serve as primary sampling unit and the enumeration area (EA)[4] as secondary sampling unit.

[2] A master sample is a sample from which subsamples can be selected to serve the needs of more than one survey or survey round (see United Nations, 2005 and 1986).
[3] The barangay (formerly called barrio) is the smallest administrative division in the Philippines.
[4] An enumeration area is composed of one or more adjacent blocks. Usually, enumeration areas are used for census data collection.

4.74 The sample households are confined to a single EA within each sample primary sampling unit. The EA was introduced as an extra stage of sampling in order to reduce travel time for interviewers. These subsamples were further classified into four independent replicates (replicates 1, 2, 3, and 4) where a replicate possesses the properties of the full master sample.

4.75 Primary and secondary sampling units were both selected with probability proportional to size (PPS), where size was the number of households based on the Census count. In each sample EA, sample households were selected with equal probability. About 5,000 sample households are covered in the survey, equally allocating at 2,500 households each for the National Capital Region (NCR) and for areas outside the NCR.

4.76 Given the allocated sample size per region, a minimum of two (2) provinces are selected randomly per region. In each province, primary sampling units are selected using Simple Random Sampling without replacement. The selection of the number of provinces and primary sampling units is completed when the allocated sample size is reached.

4.77 Table 4.1 provides a quick overview of the sample designs presented in this chapter and their main combinations suitable for BTS and CTS. Each country can use a sample design based on the availability of resources.

Table 4.1: Sample designs for BTS and CTS

Sample Designs		*Probabilistic Techniques*				*Deterministic Techniques*	
		Simple Random Sampling (without replacement)	Equal probability of selection method (EPSEM)	Probability proportional to size (PPS)	Systematic sampling	Quota sample	Random Digit Dialing (RDD)
Panel Sample (widely used for BTS)	Stratified Random Sampling	BTS CTS	CTS	BTS CTS	BTS	CTS	CTS
	Stage Sampling (with stratification in the stages)	CTS	CTS	CTS		CTS	CTS
	Cluster (combined with multi-stage	CTS	CTS	CTS		CTS	CTS

sampling)	

Chapter 5: Estimation procedures and accuracy

A. Introduction

5.1 This chapter presents commonly used methods 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 may arise with the sampling, e.g., when the sample misrepresents the target population. 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.

5.2 While the treatment of non-response is discussed in Chapter 7, the present chapter focuses on the first three types of weights. In particular, this chapter 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 an 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 (in most of the cases, percentages of reply options), 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:

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

where:

- 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). The interest is in the percentages of positive responses, neutral responses and negative responses. Thus these questions generate three binary target variables each of which can take value 0 or 1 depending on the response chosen. In the case, for example, of the percentage of positive response, the target variable y_i takes values +1 if the response of unit *i* is 'positive' and 0 otherwise. Similar is the case for the neutral and negative answers. 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 such questions 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. 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 (2) can also be used for CTS, by choosing $w_i = 1$ for all units. To avoid unnecessary duplication, size weights are included in all formulas in the rest of the chapter. If the interest is only in CTS, the weight w_i can be simply removed and the total weight W be replaced with N in all the formulas.

1. 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. However, 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.

2. 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. However, 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 economic unit in weighting all the survey answers.

5.11 If a single variable needs to be used as a 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 detailed levels (e.g. *Manufacture of footwear*, ISIC Rev. 4 Class: 1520), and the other variable, called the outer weight[1], is used when aggregating to higher levels (e.g. *Manufacture of leather and related products*, ISIC Rev. 4 Division: 15 or *Total manufacturing*, ISIC Rev. 4 Section: C).

5.12 Since value added is the best choice as a weighting variable, it should be used as the 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 classes the number of employees and value added are

[1] The outer weights (e.g. GDP) need to be updated to reflect changes of the economy. In European countries, where the economic variables, such as GDP, are quite stable, the common practice is to change weights every five years. Each country may define the frequency of updating on the basis of the changes of its economy.

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:

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

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

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.

5.14 It is easily shown that this two-step process is equivalent to using the following weights in equation (2):

$$w_i = e_i \frac{V_j}{E_j} \tag{5}$$

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.

5.15 In order to simplify the presentation, 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 (5) can be used to derive the size weights w_i from the inner and outer weights.

C. Estimators and sample weights

1. Horvitz-Thompson and ratio estimators

5.16 To estimate the population target variable based on the sample response, the standard and widely-used *Horvitz-Thompson (HT) estimator* can be used (for additional reading, see, for example, Cochran (1977), Särndal et al. (1993), and Banning et al. (2012)):

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

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

- π_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 π_i, and hence the sample weights, depend on the sample design. In Section C.2 formulas for π_i are presented for the sample designs discussed in the previous chapter.

5.17 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 $\hat{W}^{HT}$:

$$\hat{Y}^R = \frac{1}{\hat{W}^{HT}} \sum_{i \in p} \frac{w_i}{\pi_i} y_i \tag{7}$$

$$\text{with } \hat{W}^{HT} = \sum_{i \in p} \frac{w_i}{\pi_i} \tag{8}$$

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

5.18 Both the HT estimator and the ratio estimator have their pros and cons. The HT estimator is unbiased[2], which means that the estimate is, on average, equal to the target variable, whereas the ratio estimator is not. However, it can be proved that the bias of the ratio estimator is small if the sample size is large enough. It can usually be neglected under the "rule of 30" (see chapter 4).

5.19 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.

5.20 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

[2] 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.

adding a question to the questionnaire about the variable used as size weight (e.g., number of employees).

5.21 Note that formulas (6), (7) and (8) are only valid for the entire population. If interest is 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;
- $\hat{W}_Q^{HT}$ 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.

2. Inclusion probabilities

5.22 In order to use the estimators introduced in the previous section, one must know the inclusion probability for each unit in the population. This section discusses the inclusion probabilities of the sample designs introduced in the previous chapter.

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

5.24 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.

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

$$n_h = \frac{N_h}{N} n \tag{11}$$

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.

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

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

$$n_h = \frac{N_h S_h}{\sum_{h=1}^{H} N_h S_h} n \tag{12}$$

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

Where 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.

5.28 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:

$$\pi_i = \frac{w_i}{W'} n' \tag{14}$$

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 $W' = W - \sum_{i \in s} w_i$.

5.29 If one substitutes (14) in formula (6) or (7), 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.

5.30 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:

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

where:

- c denotes the cluster in which unit i resides;

- $\pi_c^{(1)}$ denotes the probability with which cluster c 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 c in the second stage of the sampling process.

5.31 Any combination of sample designs discussed in this section (including the two-stage sample design itself that gives a design with three or more stages) can be used to construct a two-stage sample design. The number of possible two-stage designs is therefore very large, but we limit ourselves to discussion of a single, realistic example.

5.32 Consider the case for CTS, where 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 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 section, we find:

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

where

- 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.

5.33 From this, the overall inclusion probability to select person i can be determined as follows:

$$\pi_i = \frac{n_h}{F_h f_c} \tag{17}$$

5.34 Comparing (17) 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.

5.35 Table 5.1 summarises the inclusion probabilities for all sample designs discussed in this section:

Table 5.1: 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	$\pi_i = \pi_c^{(1)} \pi_{i,c}^{(2)}$

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

D. Post-stratification

5.36 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. Such inaccuracies may occur 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 these cases, one may find that some categories are over- or under-represented, simply because of the random nature of the samples used.

5.37 The data collection mode can also contribute to misrepresentation of the sample. 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 more limited.

5.38 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 this end, 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:

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

$$\text{with } \hat{W}_j^{HT} = \sum_{i \in p_j} \frac{w_i}{\pi_i} \tag{19}$$

where the notation is the same as used before, with a few 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.

5.39 The ratios $W_j / \hat{W}_j^{HT}$ act as weights in equation (18); 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.

5.40 To obtain more insight into this abstract formula, we present the post-stratification estimator for the simplest case, namely the SRS without replacement while ignoring size weights. Then equation (18) simplifies considerably:

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

where the notation is the same as before, with the following additions:

- n_j denotes the number of elements in p_j;
- N_j denotes the number of elements in P_j.

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

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

5.42 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.

5.43 Other approaches must be applied when, as is often the case in practice, 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. These methods are called *incomplete post-stratification methods*; the most widely used are listed below:

- 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; and
- The Calibration method, which can be regarded as a more general method, with ratio, regression, and raking ratio estimators as special cases.

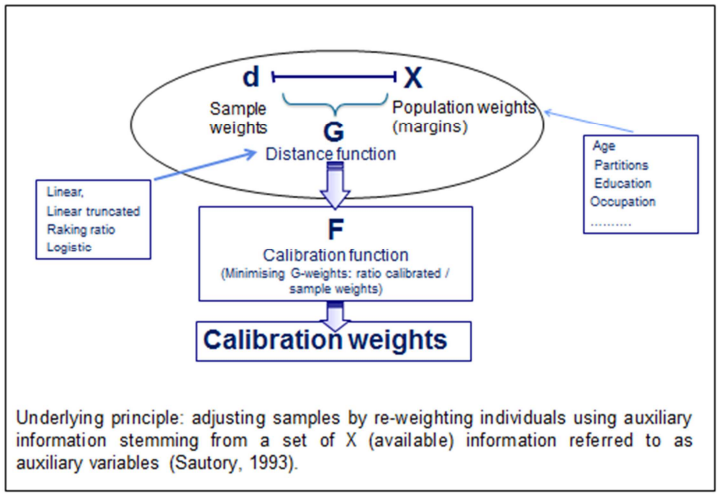
5.44 The principle underlying the Calibration method, proposed by Deville and Särndal (1992) (see also Sautory, 1993), is to adjust samples through modifying unit weights using auxiliary information stemming from a set of available information in the population referred to as calibration variables.

5.45 The adjustment is made by replacing the original sample weights with new weights as similar as possible to the original ones so that:

- For a categorical (qualitative) variable, the proportions values of the reply options estimated within the sample, after re-weighting, will become the same as 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, lowering the bias due to unit non-response (see also Chapter 7). Figure 5.1 provides a graphical representation of the method.

Figure 5.1: Graphical representation of the Calibration method



Source: prepared by ISTAT based on Fullone and Martelli (2003)

5.46 Given a population U of N individuals from which a sample p of size n has been selected and Y being a variable of interest, which needs to be estimated for the total in the population, the usual Horvitz-Thompson estimator for these purposes can be formulated in a slightly different manner as follows:

$$\hat{Y}^{HT} = \sum_{i \in p} \frac{1}{\pi_i} y_i = \sum_{i \in p} d_i y_i$$

where d_i stands for the sample weight (the reciprocal of the inclusion probability).

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

$$\hat{Y}^{CAL} = \sum_{i \in p} v_i y_i \qquad (22)$$

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

- v_i denotes the calibration weight of unit i, which takes into account both the sample weights and the values of selected auxiliary population variables X to which the sample estimate is adjusted.

5.48 The calculation of v_i is achieved by solving a rather complex non-linear optimisation problem, of which only the general principles are presented here. A detailed explanation of the method is beyond the purpose of this handbook; for a fuller explanation the reader can refer to literature (Deville and Särndal (1992), Deville et al. (1993)); or for more practical guides: Sautory (1993) and Vanderhoeft (2001).

5.49 Let $X_1, \dots, X_j, \dots, X_J$ be J auxiliary variables, available in the sample, and for which the population totals values are known:$X_j = \sum_{i \in U} x_{ij}$ where U denotes the population. These variables are referred to as calibration variables.

5.50 The calibration weights v_i, as similar as possible to the original sample weights, should verify the calibration constrains, that is, they have to :

$$\sum_i v_i x_{i,j} = X_j \qquad \forall j = 1, \dots, J \tag{23}$$

5.51 For this purpose a "distance function" $G(r_i)$ with $r_i = v_i/d_i$ is introduced to measure the distance between the g-weights and 1. If a g-weight r_i equals 1 it means that no correction for the initial weight is needed for unit i.

5.52 The inverse of the derivative of the distance function is called calibration function $\mathrm{F}(.) = \mathrm{G}'^{-1}(.)$ which allows one to calculate the calibration weights v_i.

5.53 According to Sautory (1993) four types of distance functions could be considered, each of them having particular properties of *G* and *F*, namely: (a) linear function with quadratic distance function and linear calibration function; (b) raking ratio or multiplicative function with

exponential calibration function; (c) truncated linear function with quadratic distance function and linear calibration function; (d) logit function with logistic calibration function.

5.54 Functions (c) and (d) require the setting of lower (L) and upper (U) bounds, that is the definition of the interval inside which the ratio $r_i = v_i/d_i$ can range between the calibrated and the initial sample weights. The selected calibration function determines the adjustments to be made by the calibration function. A drawback of function (a) is that calibrated weights can be negative; this can be avoided by restricting the g-weights as in (c), which imposes bounds. Truncated linear (c) and logit (d) functions are very similar within the selected (L,U) interval, while raking ratio tend to shift the g-weights upwards in a systematic way.

5.55 Returning to the calibration problem, the solution which minimises the distance:

$$\underset{v_i}{Min} \sum_{i \in p} d_i G\left(v_i / d_i\right) \tag{24}$$

subject to J calibration constraints according to formula (23) and, depending on the selected distance function, and also sometimes to boundary constraints, is obtained with a mathematical iterative process, requiring the solution of the non-linear system of J equations in J unknowns which allows one to achieve convergence when the calculated g-weight ratio v_i/d_i between two subsequent iterations becomes approximatively stable. To this aim the introduction of a vector λ of Lagrange multipliers associated with the constraints and the application of the Newton iterative method are required, which permits solution of the equation:

$$v_i = d_i F\left(x_i^{'} \lambda\right) \tag{25}$$

5.56 It is important to point out that even if the method and calculations are rather complex, some programmes are freely available on the internet for processing the calibration weights so that relatively inexperienced researchers can also apply these techniques. See, for example,

CALage sur MARges (CALibration on MARgines - CALMAR) of INSEE[3] and RGENESEES of ISTAT[4]. Box 5.1 provides a brief description of the application of CALMAR procedure in the Italian consumer tendency survey.

Box 5.1: CALMAR in practice in the Italian CTS

The application of CALMAR procedure in the Italian CTS is described below. In order to shorten the iterative process of convergence, CALMAR requires, as starting weights, the effective sampling weights, which are calculated according to formula (17). As calibration variables, besides the regional ones, those that are usually more affected by the sampling design are considered, namely: Age (4 classes), NUTS2 (19 Italian Regions), LAU2 (7 classes of municipalities according to size), Occupation (4 categories); Education (3 categories). Gender was excluded, as it was already considered in the quota interviews. The aim was to balance as much detail as possible with a reasonable convergence over time, as increasing the number of calibration variables may conflict with convergence purposes. CALMAR, in fact, foresees up to 15 iterations to achieve convergence.

Even if the survey is conducted monthly, the calibration variables are calculated and updated yearly (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, that is only marginal distribution of the selected population weights is considered. As the values of the calibration variables are calculated in percentages, the yearly total of the population is also required by the programme.

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 (defined by a lower L and an upper

[3] 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

[4] 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

bound U) of the function defined. This setting is rather subjective. Firstly a procedure was set up that, for every month, first applied the raking ratio method. The resulting minimum and maximums bounds, identified by raking ratio were used as starting point for setting the domain of the logit method, gradually decreasing the upper (and increasing the lower) bound t times as to reach values for which higher convergence (that is a difference less than a predetermined small quantity ε) between initial d_i and final weights v_i was no longer reachable. Then the $t-1$ iteration is considered the desired one. At this point the Calmar programme calculates the v_i weights for all (non null) units of the sample.

In the Italian experience the lower (L) and upper (U) bounds were, in the whole time span considered, on average, L=0.41 and U =1.55. These bounds were rather constant over time. In the period 1995-1997 their averages were L=0.33 and U=1.51. It may be possible that in this time period, the less efficient sample design was being compensated for the collapsing of the regional margins into partitions. In the 2004-2006 period, the bounds were L=0.44 and U=1.56. The final calibrated weights thus ranged from about 2,500 to 37,000, which is 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 40,000 individuals. The resulting weights, once divided by their average, are associated with each interview, in this way summing up *n*=2,000 (the sample size).

E. Accuracy and sample size

5.57 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, due to the use of sample data, between the sample estimate and the population true value is defined as the *sampling error*. The sampling error is a measure of the precision of estimates and it concerns the probabilistic nature of the sample design, which occurs when two different samples, drawn using the same sample design, do not give similar results.

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

1. General theory

5.59 The 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, α, of making an error. That is:

$$\Pr(|\hat{Y} - Y| \leq d) = 1 - \alpha \tag{26}$$

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

5.60 If the sample is large enough $\hat{Y}$ has an approximately normal distribution with variance $\text{var}(\hat{Y})$. If it is assumed that any bias in the estimate is negligible, the normal distribution is centred on Y, thus:

$$Z = \frac{\hat{Y} - Y}{\sqrt{\text{var}(\hat{Y})}} \tag{27}$$

Z has a standard normal distribution, and

$$\Pr(|\hat{Y} - Y| \leq d) = \Pr\left(|Z| \leq \frac{d}{\sqrt{\text{var}(\hat{Y})}}\right) = 1 - \alpha \tag{28}$$

thus:

$$\frac{d}{\sqrt{\text{var}(\hat{Y})}} = z_{\alpha/2} \tag{29}$$

with $z_{\alpha/2}$ the abscissa that cuts off an area of size $\alpha/2$ from the right tail of the standard normal distribution (obtained from the tables 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 $1 - \alpha$.

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

$$\text{var}(\hat{Y}^{HT}) = \frac{1}{W^2}\sum_{i\in P}\sum_{j\in P}\frac{\pi_{ij} - \pi_i\pi_j}{\pi_i\pi_j} w_i w_j y_i y_j \tag{30}$$

that is, 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:

- π_{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".

5.62 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:

$$\hat{\text{var}}(\hat{Y}^{HT}) = \frac{1}{W^2}\sum_{i\in p}\sum_{j\in p}\frac{\pi_{ij} - \pi_i\pi_j}{\pi_{ij}\pi_i\pi_j} w_i w_j y_i y_j \tag{31}$$

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.

5.63 For SRS, simple expressions for π_{ij} can be derived:

$$\pi_{ii} = \frac{n}{N}, \text{ and } \pi_{ij} = \frac{n(n-1)}{N(N-1)} \text{ for } i \neq j \tag{32}$$

5.64 For most other sample designs the exact expressions for π_{ij} are difficult to derive, making it hard to use variance formulas (30) and (31). Reasonable approximations for the variances are available in certain situations, but it would go beyond the scope of this Handbook to give these expressions for π_{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).

5.65 For the other estimators presented in this Handbook, approximate formulas for 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 it 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.

2. Example: Simple Random Sampling

5.66 While the formulas for the sample variance given in the previous section are valid for all sample designs, they do not offer much insight into the accuracy of particular sampling strategies. However, in the case of a Simple Random Sample, formula (30) simplifies considerably:

$$\text{var}(\hat{Y}^{HT}) = \frac{N}{W^2}\frac{N-n}{n}S^2 \tag{33}$$

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

$$S^2 = \frac{1}{N-1}\sum_{i\in P}\left(w_i y_i - \frac{1}{N}\sum_{j\in P} w_i y_i\right)^2 \tag{34}$$

5.67 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.

5.68 Equation (33) shows that the variance of the estimator increases with increasing values of 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 the intuition that if all units in the frame were to be surveyed (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 was the

same for each unit (in which case $S^2 = 0$), only a single unit need to be surveyed to know the value of this variable for all units, and hence obtain the exact value of the population target variable.

5.69 If the size weights are ignored and the fact that y_i can only attain the values 0, and +1 is taken into account (that is in the case of ETS categorical data), a very simple expression for S^2 can be derived:

$$S^2 = \frac{N}{N-1} Y(1-Y) \tag{35}$$

By combining this with (29) and (33), the relationship between d, the precision of the estimate, and n, the sample size can be obtained:

$$\frac{d^2}{z_{\alpha/2}^2} = \frac{N-n}{N-1} \frac{Y(1-Y)}{n} \tag{36}$$

In this formula, 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, which means:

$$d = z_{\alpha/2} \sqrt{\frac{Y(1-Y)}{n}} \tag{37}$$

or, solving for n,

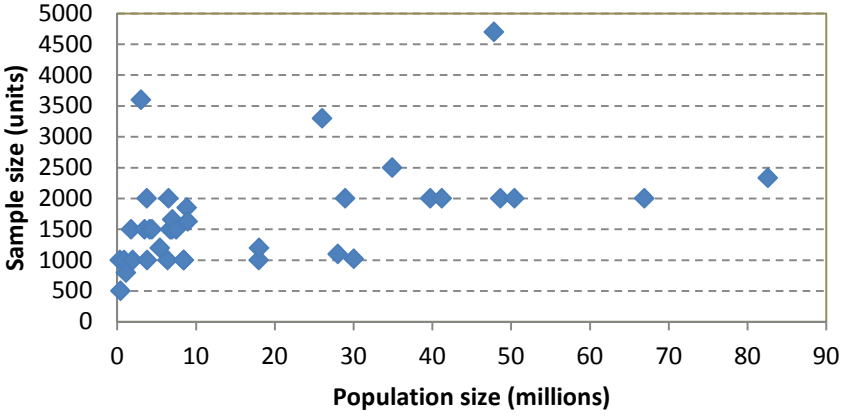
$$n_0 = \frac{z_{\alpha/2}^2}{d^2} Y(1-Y) \tag{38}$$

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

$$n = \frac{n_0}{1 + (n_0 - 1)/N} \tag{39}$$

5.70 An important consequence of (38) 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 CTS select a sample size ranging from 1,000 to 2,000 units, from populations running in the millions, as illustrated in Figure 5.2. Similar sample sizes are used in BTS (industry only), as can be seen from Figure 5.3.

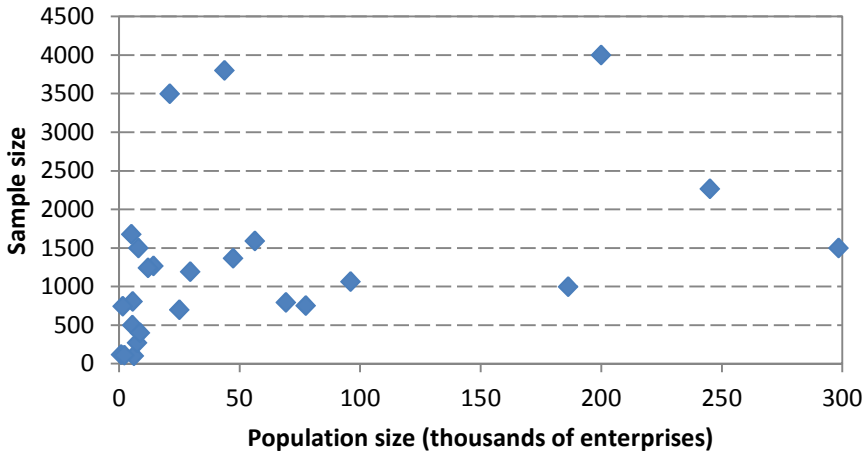
Figure 5.2: Sample size of CTS as a function of population size for different OECD, EU and EU accessing countries



Source: prepared by Istat based on EC Metadata and OECD MEI

Note: USA over 250 million population >15 years; 500 sample size; Switzerland not included

Figure 5.3: Sample sizes of BTS (industry only) as a function of population size for EU countries



Source: prepared by CBS based on EC metadata. Romania not included.

5.71 For example, if the acceptable margin of error for our estimates is $d = 0.03$ and $\alpha = 0.05$, which means a risk of 5 per cent that the actual error is larger than the margin of error, the value of n_0 is obtained as:

$$n_0 \approx \frac{1.96^2}{0.03^2} 0.5 * 0.5 \approx 1{,}067 \tag{40}$$

where $z_{\alpha/2} \approx 1.96$ and, in absence of any further information on Y, it is assumed the situation of maximum uncertainty, that is $Y = 1 - Y = 0.5$.

5.72 As the population in CTS is generally large (say $N > 100{,}000$) the finite population correction does not give any significant contribution. In fact, using (39), the result remains nearly the same:

$$n \approx \frac{1{,}067}{1 + (1{,}067 - 1)/100{,}000} \approx 1{,}056 \tag{41}$$

5.73 In BTS, however, the 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 (39) gives the following result:

$$n \approx \frac{1{,}067}{1+(1{,}067-1)/8{,}000} \approx 942 \tag{42}$$

5.74 Table 5.2 shows the outcomes for simple random samples according to formula (38) for different values of Y and different desired precisions. It is worth noticing that, d being constant, the sample size does not vary dramatically in the interval between 0.3 and 0.7. On the other hand, the sample size heavily depends on d, for example, if the precision is increased from $d = 3\%$ to $d = 2\%$, the sample size has to increase by more than double; increasing the precision from $d = 2\%$ to $d = 1\%$, the sample size needs to be increased by nearly four times.

Table 5.2: 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	3.457	6.146	8.067	9.220	9.604	9.220	8.067	6.146	3.457
0.02	864	1.537	2.017	2.305	2.401	2.305	2.017	1.537	864
0.03	384	683	896	1.024	**1.067**	1.024	896	683	384
0.04	216	384	504	576	600	576	504	384	216
0.05	138	246	323	369	384	369	323	246	138
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

5.75 The cost of the survey is a further important factor that has to be considered. The obtained sample size may be too costly for the available budget and it may be necessary to accept a larger margin of error (i.e., by increasing d). In that case equation (37) can be used to estimate the level of precision that can be attained. Table 5.3 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 5.3: 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
1,000	1.9	2.5	2.8	3.0	3.1	3.0	2.8	2.5	1.9
1,250	1.7	2.2	2.5	2.7	2.8	2.7	2.5	2.2	1.7
1,500	1.5	2.0	2.3	2.5	2.5	2.5	2.3	2.0	1.5
2,000	1.3	1.8	2.0	2.1	2.2	2.1	2.0	1.8	1.3
2,500	1.2	1.6	1.8	1.9	2.0	1.9	1.8	1.6	1.2
3,000	1.1	1.4	1.6	1.8	1.8	1.8	1.6	1.4	1.1

3. Design effect

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

$$Deff = \frac{\text{var}(\hat{Y})}{\text{var}(\hat{Y}_{SRS}^{HT})} \tag{43}$$

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

5.77 It would go beyond the scope of this handbook to discuss the design effect for all sample designs introduced in this handbook. The focus here is only on two important examples; the

interested reader is referred to the literature (for example, Kish, 1965 chap. 5.4; 8.2; 14; Banning et al. (2012) for a more accessible introduction).

5.78 Kish verified that, comparing the variances of estimates obtained from a simple random sample, a stratified random sample with proportional and Neyman allocation, the following can be obtained:

$$\text{var}(\hat{Y}_{NEY}^{HT}) \leq \text{var}(\hat{Y}_{PROP}^{HT}) \leq \text{var}(\hat{Y}_{SRS}^{HT}) \quad (44)$$

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

5.79 It is worth remembering, however, that if the variances of the target variable are very low, as is the case of tendency surveys, which deal with proportions, even the most efficient complex sample designs offer only a moderate increase in efficiency compared to the simple random sample.

5.80 As an example, in the Italian CTS, with a stratified random 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.

5.81 When dealing with clustered sample designs, where variance concentrates between clusters, instead, *Deff* may assume values larger than 1.

5.82 In these cases the measure utilized is the intraclass coefficient of correlation ρ , which, when written as roh can be remembered as the rate of homogeneity of the sample and it is given by:

$$roh = \frac{Deff-1}{B-1} \quad (45)$$

where B is the number elements per cluster so that $a * B = n\ sample\ elements$ and with a = *number of clusters* selected from the population.

Box 5.2: An example of estimation procedure (Philippines)

BTS

In a first step, differences between positive percentages of reply options and negative ones are computed for a given indicator at division level (the sample is stratified by geographic regions and divisions according to the Philippine Standard Industrial Classification (PSIC) – for details see chapter 4), with the following formula:

$$CI^{1}_{h} = \left(\frac{n_{hp}}{n_h}\ x\ 100\right) - \left(\frac{n_{hn}}{n_h}\ x\ 100\right)$$

where CI_h = difference in the division h

n_{hp} = number of firms with positive/improving/up responses in the division h

n_{hn} = number of firms with negative/deteriorating/down responses in the division h

n_h = total number of firms in the division h

The CIs for the four kinds of economic activities (industry, construction, services, and wholesale and retail trade – Sections of PSIC) are computed using the following equation:

$$CI_k = \frac{\sum_{h=1}^{s} w_{h} CI_h}{\sum_{h=1}^{s} w_h}$$

where CI_k = difference at section level

CI_h = difference at division level

w_h = N_h/N where N_h is the number of firms in the population, for the division h, and N is the population size

h = 1, 2, 3 …..s (no. of divisions per sections)

The overall indicator is computed as follows: $CI_{overall} = \sum_{k=1}^{4} w_k CI_k$ where CI_k = difference for section and w_k = weight for section (number of firms).

CTS

Weighted proportion (in percentage) for each reply option is computed. In a second step, differences from the computed proportions are calculated by subtracting the percentage of negative replies (which is the percentage of respondents saying "better", or its equivalent) from percentage of positive replies (those saying "worse", or its equivalent). Differences are calculated for categories.

Concerning the weights, base weights are provided by the Philippine Statistics Authority along with the list of samples. Adjustment of base weights after the survey is also computed. The initial step in the construction of weights is to determine the unit's base weight, which is in general the inverse of its selection probabilities. The base weight is further adjusted to take into account possible non-response and possibly to make the estimates conform to some known population totals.

Regarding non-response, which can be unit or total non-response, adjustments are made to the base weights to compensate for non-response by sampled units eligible for the survey. In essence, the adjustment inflates the base weights of "similar" responding units to compensate for each non-respondent. The most common form of non-response weighting adjustment is a weighting class adjustment that is the type of adjustment being used for surveys based on the 2003 master sample. The full sample of respondents and non-respondents is divided into a number of weighting classes or cells and non-response adjustment factors are computed for each cell.

The final survey weight assigned to each responding unit is computed as the product of the base weight, the non-response adjustment, and the population weighting adjustment. The final weights are used in all analyses to produce valid estimates of population parameters.

Chapter 6: Data collection

A. Introduction

6.1 Data collection involves the 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, especially for data collections which are conducted on a regular frequency, such as monthly ETS.

6.2 The data collection strategy strongly depends on the available financial and manpower resources of the implementing institution. Designing a data collection strategy may also depend on other key elements, such as the variables to be collected or how the questions are 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. A good strategy helps in structuring the process of data collection, minimizing measurement and non-response errors and maximizing the willingness to cooperate of respondents and thus the response rate. It also makes data collection cost efficient.

6.4 This chapter reviews different data collection strategies and survey techniques that are used for tendency surveys. In particular, Section B provides an overview of the main elements of a data collection strategy and the main factors that affect the choice of data collection. Section C describes the process of data collection from the preparation prior to conducting the survey, the data collection modes and the measures that can be taken to enhance the response rate. Finally Section C provides a quick summary of the tasks in the organizations of the data collection.

B. Data collection: designing a strategy

1. Survey techniques and communication with respondents

6.5 A data collection strategy is characterized by two important elements; the technique used to send and receive the information ('mode') and communication with the respondent. 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, mixed-mode approach, is sometimes used to optimise the data collection (see Section B.2).

6.6 The second aspect of a data collection strategy is the communication with respondents. Good communication helps to convince the respondent to participate, to eliminate misinterpretation of the request for information and possibly to counter problems in receiving the information. Communication with respondents occurs during pre-survey stage, during field work stage or in between surveys. These aspects will be described in the next sections.

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).

2. 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 the survey is about and which population is being researched. Second, quality aspects play an important role; the different sources of errors must be limited. Finally, organisational aspects also have an impact on the data collection strategy. These three factors are described below.

Scope of the survey

6.9 ETS can be conducted among either businesses or consumers. For data collection, this has consequences. Usually, businesses are more complex to observe, due to their heterogeneity. It can be more difficult to find the right contact person for business as compared to 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 particularly 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 an 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 answerable intuitively and the questionnaire is fairly short. It should be mentioned that, although the questionnaire and the mode of data collection influence each other, once the data collection mode is selected, the questionnaire has to be aligned to suit the selected mode.

Quality aspects

6.12 Data collection is affected by measurements errors. Measurement error is the difference between a measured value of quantity and its true value (Dodge, 2003). These errors can occur in the recording of the true answers or in the answers themselves.

6.13 Four primary sources of measurement error are distinguished (UNSD, 2005):

- *Questionnaire*: measurement errors related to the questionnaire design, such as its visual layout, the topics it covers, and the wording of the questions. In ETS these measurement errors are greatly reduced by using the internationally agreed questionnaires.
- *Data-collection method*: measurement errors related to how the questionnaire is administered to the respondent (for example, mail, in person, or diary). Respondents may answer the questions differently in the presence of an interviewer, by themselves, or by using a diary.
- *Interviewer*: measurement errors related to the interviewer. The interviewer may introduce error in the survey responses by not reading the items as intended, by probing inappropriately when dealing with inadequate responses, or by adding other information that may confuse or mislead the respondent.
- *Respondent*: measurement errors related to the respondents, who, because of their different experiences, knowledge and attitudes, may interpret the meaning of questionnaire items differently.

6.14 There are a number of approaches to quantify measurement error such as randomized experiments, cognitive research studies, repeated measurement studies, and record check studies (see UNSD 2005 for a detailed presentation of such approaches). In general, quantifying the existence and magnitude of a specific type of measurement error requires advance planning and thoughtful consideration. Unless small-scale (that is to say, limited sample) studies are

conducted, special studies are necessary that require randomization of subsamples, re-interviews, and record checks. These studies are usually expensive to conduct and require a statistician for the data analysis. Nevertheless, if there is sufficient concern that the issue may not be adequately resolved during survey preparations or if the source of error is particularly egregious in the survey being conducted, survey managers should take steps to design special studies to quantify the principal or problematic source of error. (UN, 2005).

6.15 Another kind of errors stems 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 electronic form, processing errors can be dramatically reduced. The questions are automatically posed in the right sequence and the answers are verified immediately. Checks can be built in to correct answers that are outside of the acceptable range. Data quality controls, such as automatic checks or double data capturing, should be put in place to mitigate data capturing errors.

Organizational aspects

6.16 There are some considerations on the organizational aspects that should be taken into account when designing a successful data collection strategy for ETS. They include the frequency and timeliness of ETS, the financial costs of the survey and the response burden. Ultimately the data collection strategy is a trade-off between the benefits and costs. One unique feature of ETS is its timeliness. The results are often published very shortly after the end of the reference period (mostly monthly). The total processing period is generally limited to a maximum of 20 working days. The data collection period is often limited to the first two to three weeks of each month. This very short collection period limits the use of some survey techniques.

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 always the pressure to reduce the response burden on respondents (especially businesses). A data collection strategy thus should make it easy for the respondents to comply and not over-burden the respondents. Questionnaires must therefore be as user-friendly and as 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.

C. Designing the process of data collection

1. Pre survey: preparations and communication

6.19 Preparations and communication with the target respondents prior to the deployment of the survey is an 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 are usually drawn from available lists 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 carried out 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, initial coordination is not necessary.

2. Data collection modes

6.22 There are different ways to collect data in a survey. The first is whether the respondent is assisted by an interviewer or completes the questionnaire on his own. The second distinction is the form in which the questionnaire is presented and recorded.

6.23 A personal face to face interview assisted by a computer is referred to as Computer Assisted Personal Interviewing (CAPI). The 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 can take place over the telephone and is referred to as Computer-Assisted Telephone Interviewing (CATI). This form of data collection is mostly organised from an internal or external call centre. The interviewers can be centrally instructed, guided and supervised. The costs are relatively high, although not as much as compared to CAPI,

and the interviews can be done fast. This form of interview is less suited for long and complex questionnaires. In general, this is quite an effective mode for CTS.

6.25 Telephone and face to face interviews generally avoid some forms of measurement errors and the issue of non-response. With these modes, the required information is immediately and fully available. The interviewers usually ensure 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. However, although these modes lead to more complete responses and higher response rates, they tend to be more costly - because of the required staff- and they require a longer period of time for the field work as compared to other collection modes. In addition, there is a larger burden on respondents as interviewers go to the respondents' homes or companies and the visits can be time consuming.

6.26 For ETS, it is essential that the data collection period be short. The economic conditions can change every day, thus influencing the survey results especially in cases 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.

6.27 An alternative to assisted interviews is the traditional way of conducting surveys with paper questionnaires. The respondent receives the questionnaire by mail, completes it without help of an interviewer and sends it back by mail. The disadvantage of this type of observation is that the back and forth sending of postal questionnaires leads to some delay. An alternative is using e-mail for sending and returning the questionnaire. In this case the delay in sending and receiving the forms is much reduced as well as the associated costs (stamps, envelopes etc.). Both postal and e-mail questionnaires pose a risk of incomplete answers or partial / item non-response. Also, questionnaire by regular mail usually yields low response rates.

6.28 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, 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 could be implemented in the form of hard checks that would prevent the respondent from continuing until a probable answer has been given and soft checks that would only display a warning to indicate to the respondent an improbable answer. Data analysts generally prefer questionnaires with automated checks as these represent a first layer of data validation. However, too many automated checks can irritate respondents and lead to non-response if they quit answering halfway through the questionnaire. Furthermore, improbable answers may actually reflect the real world.

6.29 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. 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.

6.30 The different modes of data collection each have their advantages and disadvantages. On the basis of these considerations, a choice will have to be made of the preferred mode. Table 6.1 shows the main advantages and disadvantages mentioned for each mode.

Table 6.1: Advantages and disadvantages of data collection modes

Mode		Advantages	Disadvantages
Interviewer-administered	Face to face (CAPI)	• interviewers can motivate respondents • completed questionnaires • possibility to explain and clarify questions • high quality data • permits lengthy interviews	• expensive • socially desirable answers • complexity, long fieldwork period
	Telephone (CATI)	• central instruction and supervision possible • timeliness • completed questionnaires • possibility to clarify questions	• inappropriate for lengthy interview • hasty answers • socially desirable answers • relatively expensive
Self-completed	Paper questionnaire (or questionnaire sent and received by e-mail)	• sensitive subjects can be researched • available to respondents with no internet access • sending and returning is cheap	• item non-response • limitations to complexity of the questionnaire • trouble reaching the right contact person in businesses • postal delay • data entry, processing and correcting can be costly
	Web forms	• timeliness	• access to internet is needed

Mode		Advantages	Disadvantages
		• low cost • completed questionnaires, 'hard' and 'soft' checks for improbable answers • automated processing	• low motivation of respondents • low response if survey is not mandatory
	Mixed Mode	• high response rate, reduced (selective) non-response • timeliness	• costs • mode effect

3. Response enhancing measures

6.31 There are a number of measures that can be implemented to enhance and facilitate the responses to a survey, thus increasing the overall response rate. Response enhancing measures can be divided in two groups: *proactive measures*, which are applied during the preparation of data collection (during the development of the data collection strategy) and *reactive measures* which are applied during the data collection phase when the response is unsatisfactory.

6.32 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.

6.33 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 measures. 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 measures

6.34 In general it is a good practice to keep in mind the point of view of the respondent when designing a survey, communicating and collecting data. Completing a survey questionnaire has a low priority for most respondents as it takes time with no evident gains in return. Thus, it is important to communicate to the respondents the significance of the survey, as well as of the importance of the particular respondent's participation in the study.

6.35 The quality of the material used in the survey influences the response rate considerably. A good questionnaire should be well-designed, visually attractive and should only ask for relevant information. It is usually important to provide the contact of a person from the survey group who can respond to questions/concerns on a timely basis. 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. Also, sharing the results of the survey with the respondents also provides a good incentive to participate in the study.

6.36 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 measures

6.37 Participation in a survey depends to some extent on whether completing the questionnaire is mandatory or voluntary. The response rates in voluntary surveys are generally lower than in surveys that entail a legal obligation to cooperate. A good reminder policy is especially important 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 e.g. what described by Statistics Netherlands in Box 6.1 and by Philippines in Box 6.2) 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 developed taking into consideration various

elements such as costs, speed and effectiveness. For example, if reminders are sent too quickly or too frequently, this may have a negative rather than positive impact on the respondents.

Box 6.1: An example of data collection strategies (Statistics Netherlands)

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

BTS

For the BTS, CBS 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 CBS. 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.

CBS 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 regular mail 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 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 per cent. 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, CBS has a service centre available for respondents, so that they can contact CBS for questions, exchange contact information and report their complaints, etc.

In the past, CBS 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. Having fairly little impact on the response rate, this strategy was replaced with the current approach of web form only.

CTS

The data collection mode for CTS is a personal interview by phone (CATI) and the panel changes monthly. Because of the higher response risk and given the limited size of the panel, CBS decided to interview the consumers in person, by telephone. This approach has the 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 6.2: An example of data collection strategies (Philippines)

In the Philippines, ETS are conducted by the 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 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.

D. Organizing the data collection

6.38 Before proceeding to the actual data collection, it is essential that all the necessary documents have been finalized and the required processes and systems have been undertaken and developed. A pre-test of the tasks to be undertaken and a testing of the systems are also prerequisites to having smooth implementation of data collection.

6.39 It is also important to be aware of the practical considerations and best practices for addressing any problems or concerns that may be encountered during the data collection. Regardless of the mode for gathering the data, having a well-organized data collection is a must as the quality of the survey would largely depend on the inputs to be provided by the respondents.

6.40 In organizing the data collection, the checklist below can be used as a guide. Adaptations to the specific country situation regarding data collection mode and available resources have to be made.

(a) Other documents to be prepared aside from the questionnaire, such as the letters, listing sheets with contact information, monitoring forms, field operations manual or enumerator's manual, etc.;

(b) Identification, selection and training of data collectors;

(c) Identification of supervisors and organization of teams;

(d) Training to be provided to the data collectors;

(e) Schedule of data collection, including follow-ups;

(f) Other supplies and logistics needed in data collection;

(g) Technical facilities for data capture, such as computers and programs.

6.41 There could be some variations in training requirements among the different data collection modes. To ensure consistency in results, all data collectors should be carefully trained. It is expected, however, that the training for face-to-face interview, which requires fieldwork, can be more extensive compared to the other modes. Data collectors have to know the objectives of the survey and the target respondents, and to understand all 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. Considering that the ETS questionnaire is not very complicated, it can be expected that the training required for data collectors also will not be complicated.

6.42 For face-to-face interview and telephone interview, having a field operations manual or enumerator's manual would definitely guide the data collectors. The manual articulates the information about the survey and provides instructions and explanations to the items in the questionnaire. Although the manual is not necessary for the other types of data collection, it can serve as a reference for the other data collectors. For the interview type, it is important that the questions be clearly communicated verbally to the respondents in order to expect appropriate responses. In the case of computer aided interview, such as CAPI and CATI, the required computer programs should be developed in order to facilitate the data collection.

6.43 For paper questionnaires and web forms, it is important to provide additional instructions or explanations for clarity. For emailed questionnaires and web forms, it is critical that the required computer programs or systems to be used for the data collection be fully functional and have been extensively tested.

Chapter 7: Managing sources of non-sampling errors

A. Non-sampling errors

7.1 The total error which affects estimates obtained from a sample survey can be due to *sampling* errors and *non-sampling* errors. Sampling errors arise because of the sampling itself whereby the estimate of the target variables is based on a subset of units of the whole population. Thus, the outcome from the interviews is based on a selection of units and not 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. Chapter 4 and 5 deal with the sampling methods and estimators and thus address sampling errors.

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 when entering them into the database (codification, registration, revision). Since there is not a complete and unique method able to address all the possible non-sampling errors, the best practice consists in building up the "error profile" (i.e. 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 context, 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:

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

where σ^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 σ^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 σ^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 summarized as follows:

a) Sampling error

b) Non-sampling errors

   i) Coverage

   ii) Measurement and processing

   iii) Non-response

      - Total
      - Partial

1. 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, the kind of activity, juridical status, size and location; and for households, kind of municipality, number of components); and 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.

7.7 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 more than one name in the list correspond.

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

7.8 In the case of non-completeness, the population is not completely included in the list (for example, fixed telephone lists). Consider y as the target variable, P_L as the relative frequency of units correctly included in the list L (and as $1 - P_L$ the relative frequency of those not included);

and $\bar{y}_L$ 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:

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

If $\bar{y}$ is estimated 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:

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

7.9 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:

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

7.10 In the case of *clusters of units*, if clusters are rare and small, all units belonging to the clusters can be included in the sample, in order to give to each unit the same inclusion probability as the name drawn from the list (for example, the name is an address, the units are all the households living at the same address). Efficiency is lost if clusters include too homogeneous units. Otherwise, only one unit from the cluster can be chosen at random and be assigned 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.

7.11 The case of *unknown, or not existing names or over-coverage*, occurs for example, when households moved away, or firms stopped or changed their activity. In this case, it is necessary to 1) identify not eligible units, and 2) decide on accepting or not a sample size lower than that

desired. If substitutions are selected, the rate δ of not existing names could be estimated in advance and followed by selecting a sample of size $n/(1-\delta)$ in order to achieve a sample with size n. Otherwise, if a substitution list is used, it must be created according to the same sampling scheme used for the selection of the first sample.

7.12 The case of *replicated names* is when, for example, a firm has its local units included in the list with distinct names, so that the firm is present in the list more than once.

7.13 Generally, the inclusion probability is given by the sum of all the inclusion probabilities of the corresponding labels in the list. In a self-weighting design the ratio estimator is the following:

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

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 πA_i, where π is the constant inclusion probability (under the self-weighting hypothesis) of labels. Since the probability to find the same unit in the same sample more than once is quite low, it is not convenient to limit actions to eliminate duplications from the sample.

The case of telephone surveys

7.14 When telephone consumer surveys are carried out, normally the list used for drawing the sample is the national list of households 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 problematic. 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.

7.15 Each case above leads to uncontrolled modification of the original inclusion probabilities defined through the sampling design, and may produce bias in the final estimates. In particular, under-coverage will produce a bias which increases with the differences in 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 that the sampling variance increases. A simple way to tackle the problem – as already mentioned – consists of incrementing the sample size to take into account non-eligibility of some sample units, or of using substitute lists.

7.16 With regard to the bias, it should always be 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 of 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 note 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.

7.17 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). This is based on the idea of modifying 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 non-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.

2. Measurement and processing errors[1]

7.18 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, in 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 by adopting modern tools for obtaining and saving responses (electronic questionnaires, via web or CAPI, CATI techniques[2]), which may already include some basic checks regarding measurement errors.

7.19 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

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

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, frequent contacts with respondents in CTS should preferably be avoided. These respondents are often rotated at each survey round because the "surprise" effect characterizing new entries into the sample may be a key issue in order to achieve highly reliable responses.

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

Measurement Errors

7.21 Measurement errors occur when the observed value is different from the true one. This can be due to: the behaviour of the respondent unit; the instrument used to get information (kind of questionnaire); the effect on 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 refer to the first kind only).

7.22 Potential problems concerning the responding unit can be summarized as:

(a) Lack of capability to report correctly;

(b) Insufficient knowledge to answer correctly;

(c) Lack of motivation to report correctly.

Errors due to the questionnaire may be caused by:

(d) Ambiguous wording of questions;

(e) Unclear layout of questionnaire.

7.23 Problem (a) arises, for example, when the responding unit differs from the reporting unit and it is not able to provide the information for the reporting unit. This is the case, for example, in BTS when the sample unit is the enterprise (typically the survey frame is the business register which consist of enterprises) and the enterprise manager is asked to either complete the survey for the reporting units (KAU) belonging to its enterprise or send it to them for their completion. In the first case, the enterprise manager may not be able to accurately answer the questions for all its reporting units.

7.24 As regards CTS, a similar problem is due to the 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.

7.25 Problem (b) arises because of the respondent's lack or insufficient knowledge on the questions in the survey. This 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.

7.26 Problem (c) arises when the respondent is not motivated enough to provide accurate answers to the questionnaire. The best way to minimize cause (c) 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.

7.27 Problems (d) and (e) refer to the layout and wording in the questionnaire. They 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.

7.28 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.

7.29 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 of comparing observed and true values (of course, when available). A second method consists of replication of interviews using more expert interviewers. But also in this case we cannot rule out measurement errors also occurring 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

7.30 Processing errors may be introduced during:

- data entry;
- data editing;
- data tabulation.

7.31 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.

7.32 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.

7.33 In the context of any data editing process, the main risk is that errors may be introduced by making the wrong adjustments. 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.

7.34 It is important to record the errors that are detected in the editing process. This helps to identify areas for improvement 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.

7.35 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.

7.36 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 that 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 by giving priority to large firms.

7.37 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.

3. Non-responses

7.38 Non-responses are due to refusal or inability 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 non-respondent units are different.

7.39 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 behaviour (large deviation of answers to the responses from the other respondents) is avoided. Non-response can be divided in two categories:

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

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.

7.40 Preventing unit non-response is a crucial element of data collection. During the fieldwork period, constant monitoring of the response rates is necessary. When the response rates are low, the data collection strategy provides for a strategy to improve response rates, deals with large firms, gives rules for re-interviews and possibly non-responding firms or consumers to be substituted. Chapter 6 deals with these elements of data collection.

7.41 Even if in some countries answering surveys conducted by public institution is mandatory by law, it is not advisable to force a 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 at motivating and convincing participants of the usefulness of their contributions, for example, by letter and /or with some incentives (e.g. small remunerations, small reward etc.).

7.42 However, unit and item non-response will still be present when the data collection period is finished and the data are 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); and (b) to increase the quality of the micro file and/or of the parameter estimates.

7.43 If we want to use imputation to improve quality, 'the quality of what' should be clear. Often, the primary goal is to accurately determine means and totals. The aim could be 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.

7.44 Response rates are useful tools to monitor the data collection process and are quality indicators of 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. The simplest measure of non-response, $NR1$, is the percentage of enterprises in the sample from which information for the actual survey was not obtained. This is defined as:

$$NR1 = \left(\frac{n'}{n}\right) \cdot 100 \quad (5)$$

where n is the number of enterprises in the survey; and n' is the number of enterprises which did not submit usable information.

7.45 This measure is useful for checking the efficiency of the data collection procedure. It is also a good indicator of the extent of non-response in censuses and sample surveys with uniform sampling fractions where all reporting units have equal weight.

7.46 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 $NR1$ is not a good indicator of the extent 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 ($NR2$) is then calculated as follows:

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

where f_i is the sampling probability for unit i.

7.47 If, in addition, responses are weighted by the size of the reporting units, the correct measure to use (here called $NR3$) is:

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

where w_i is the size weight for unit i.

7.48 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 per cent. Without the use of a fixed panel, however, the response rate will need to be somewhat higher – 60 per cent to 70 per cent.

7.49 Often the combination of $NR1$ with $NR2$ or $NR3$ can already provide good indication on the distribution of the collected responses across company weights. For example, when the un-weighted non-response rate ($NR1$) is lower than the weighted non-response rate ($NR2$ or $NR3$) 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

1. Total non-responses

7.50 The main problem is the possible bias due to non-respondents. It can be faced in a way similar to that seen for coverage errors[3]. If P_R is the relative frequency of respondent units (thus $1 - P_R$ is the relative frequency of those not respondent) and $\bar{y}_R$ and $\bar{y}_{NR}$ are, respectively, the mean of y in the two parts of the population, the following can be obtained:

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

7.51 Similarly to what was seen in Section A.1, if $\bar{y}$ is estimated 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:

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

7.52 It depends on the share of non-respondent units (in the population) and the difference between the means in the two subpopulations. The bias depends on the share of non-respondent units (in the population) and the difference between the means in the two subpopulations. It 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

[3] In the literature there is a distinction between deterministic and stochastic non-response. A discussion on these concepts can be found in the references provided in this chapter.

bias in the overall error could be higher relative to the sampling variance. In practice, bias can only be evaluated using sources external to the current survey.

7.53 A first tool for tackling non-responses simply consists of increasing the number of responses using call backs. But call backs are costly and must be carried out within a limited time horizon. In practice, often non-respondents are substituted with other units chosen at random.

7.54 A second well known methodology is the re-weighting of respondents. Under the assumption that propensity to respond is related to levels of the variable object of interest, the population is split in L strata, in order to work with post-strata having 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. The bias is given by:

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

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

7.55 In order to minimize the bias, the means of respondents and non-respondents should be quite similar in the same stratum and the response rates among strata should be quite different (if they are all equal, the bias in formula (11) turns out to be equal to formula (9), so that bias cannot be reduced).

7.56 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, it 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).

7.57 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, τ_{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 π_{hi}. The product $\pi_{hi}\tau_{hi}$ is the estimate of the *final response probability* (two phases sampling), so that formula (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.

2. Partial non-responses

7.58 While total non-responses are treated using re-weighting, partial non-responses are treated using imputation. The easiest way consists of 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. A few different imputation methods are described below.

7.59 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.

One can consider different auxiliary information in order to evaluate its efficiency for the imputation. For example, initially one 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.

7.60 In *deductive* or *logical imputation* one examines 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 hold true but only very probably does so.

7.61 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'.

7.62 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.

7.63 These methods are not often used in business and consumer tendency surveys. These methods can be used for imputations when the answering scheme allows for the answer 'not applicable' or 'not known'. However, for most questions, this is not the case.

7.64 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:

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

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

7.65 A more robust method than mean imputation is *median imputation*, since it is not influenced by outlying responses. It can be simply applied substituting $\bar{y}_{\text{obs}}$ with Med_{obs}, which is the median calculated on respondent units.

7.66 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.

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

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

7.67 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.

7.68 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.

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

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

7.69 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. This is based on 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.

7.70 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.

7.71 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; and

(b) Generally longitudinal data are not only assessed cross-sectionally, but also assessed by changes over time. To correctly estimate these changes, it is important that the imputation takes into account previous and future values.

7.72 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.

7.73 *Last observation carried forward* (LOCF) 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.

$$\tilde{y}_i^t = y_i^{t-1} \tag{16}$$

7.74 This method is mainly applicable to categorical variables which are known to 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 quantitative variables like months' work performed or capacity utilization rate.

7.75 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.

$$\tilde{y}_i^t = \frac{\sum_{obs} y_i^t}{\sum_{obs} y_i^{t-1}} \cdot y_i^{t-1} \tag{17}$$

When previous period data are not available, auxiliary data (x) can be used for ratio imputation, simply substituting y_i^{t-1} with x_i^t.

7.76 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.

7.77 Finally, the methods recommended for the treatment of missing data in ETS should take into consideration the following:

- the precise nature of the procedures used in the treatment of item non-response or missing data should be described in the metadata;
- As a minimum requirement, it is recommended to closely monitor the impact of missing data (especially for large firms) and develop a clear set of strategies to minimise non-response;
- Using 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.

7.78 For qualitative questions, it is recommended to assume the same distribution over the response alternatives [(+), (=) and (–)] as the responding reporting units in that industry. For questions requiring answers in percentages or numbers, assume that the non–responding reporting units have the mean value of responding reporting units in that industry.

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 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 (see Chapter 3). For multiple-choice question, the basic results are obtained in the form of three or more percentages (see Chapter 5) according to the number of reply options. For analytical purposes, it is difficult to interpret the movements over time of all the percentages (three or more) when they are presented simultaneously. For this reason, it is common to convert the replies options into a single number to obtain a time series useful in economic analyses. Quantified information extracted from these surveys, such as, the University of Michigan's Consumer Sentiment Index and consumer inflation expectations, are often used by researchers and policymakers in forecasting macroeconomic aggregates (Breitung and Schmeling, 2013) and testing canonical economic theories like the rational expectations hypothesis (REH) (Lahiri, K. and Zhao, Y., 2013).

8.2 Therefore, a relevant aspect in this type of surveys is the quantification of the answers, that is, conversion into a time series. While balances and diffusion indices are commonly used in reporting the results of almost all ETS, there is a considerable body of research into alternative methods. With regard to this aspect, a very important branch of research assumes that the answers to qualitative surveys are drawn from various types of probability distribution functions. 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). Given that the performance of alternative measures[1] and balances are, generally speaking, very similar, it is worth noting that the choice of a method depends on the use of the results. In the next sections, a general exposition and review of quantification methods, including the latest proposals, are described.

1. 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 are defined as 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 positive, neutral and negative option, 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 may be calculated as:

$$B = (PP+1/2 \times P)-(1/2 \times N + NN)$$ [2]

8.5 According to these formulas, 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

[1] See D'Elia (1991) for a comparison between measures using probability distribution functions and balances.
[2] This is the formula commonly used at European level (e.g. European Commission).

respondents choose the positive (or the most positive) option. The formulas described above are the most common used for the calculation of the balance.

8.6 In the case of five reply options, other weights may be used for the calculation of the balance. For example, in Italy the balance for CTS is currently calculated as B=(2 × PP+P)-(N+2 × NN). In this case, the balance varies between -200 and 200. In order to obtain always positive values for the balances, a standard procedure is to add, to each balance, the max possible value. This occurrence may be useful in calculating Confidence Indicators (see chapter 10). For instance, for the calculation of Consumer Sentiment Index by the University of Michigan 100 is added to the difference between the percentage giving favourable replies and the percentage giving unfavourable replies; thus the balance may vary from 0 to 200.

8.7 The reduction of the vector of replies to a single number, the balance, implies some loss of information. In particular, the absence of the percentage of "no change" replies, N, implies a loss of information concerning the degree of certainty of the respondents because for a given value of the balance, the certainty is negatively related to the value of N. Thus, in some cases, the balance is presented together with the percentage of replies indicating no change. Business and consumer tendency survey results are usually analysed on an aggregate basis, considering the balance. Although some loss of information occurs in calculating the balance, long experience has shown that the balance is a reliable and also practical tool for presenting survey results. It provides an easy-to-compute, easy-to-understand quantification of survey results and as such it is considered a useful tool to analyse business cycles.

2. Diffusion indices

8.8 Another common way of quantifying the information contained in the survey is the diffusion index. Generally speaking, diffusion indices 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.

8.9 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.

8.10 The Purchasing Managers' Index, PMI, calculated by Markit Group by the Financial Information Services, is an example of a composite indicator calculated by diffusion indices. For instance, for the United Kingdom (UK) manufacturing industry, Markit UK calculates a PMI composite indicator; the index is defined based on the specific characteristics of ETS in the UK and uses the following diffusion indices: 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 increase in the UK manufacturing activity, below 50 indicates an overall decrease.

8.11 The formula that links the Balance to the Diffusion index is the following:

$$DI = (100+B)/2 \text{ or } B = 2\times(DI-50)$$

where B is the balance and DI is the diffusion index[3]. 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.

[3] The conversion formula derives from: $B=100\times(P-N)$ and $DI=100\times(P+U/2)$; $P+U+N=100$, where P is Positive reply, U is Unchanged reply and N is the Negative one.

3. The probabilistic approach and other methods

8.12 Two of the main shortcomings of the balance statistics are some strict assumptions about the distribution of the answers. First, the "stay the same", U, share is constant over time and, second, the relationship between 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 exceptional events impact heavily on the correlation between the survey data and the reference 'hard' data[4].

8.13 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 the 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 $\delta^- \leq y_{it} \leq \delta^+$

[4] 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.

(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^+ and n_t^- denotes the number of respondents in t reporting an increase and decrease respectively, $p^+_t = \text{Prob}(y_{it} > \delta^+)$ and $p_t^- = \text{Prob}(y_{it} < \delta^-)$. Notice that δ is constant over time.

8.14 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 σ_t yields:

$$\sigma_t = (\delta^+ - \delta^-) / q_t^- - q_t^+ \quad (3)$$

and replacing σ 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 $z_{2t} = q_t^+ / (q_t^- - q_t^+)$.

8.15 The average μ_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[5]. The major advantage of a probability approach is that it is theoretically appealing. In addition, the results depend only on the observed percentages of answers and, only to a minor extent, on the assumptions on the probability distribution of the variables and on the indifference thresholds of respondents. However, this approach could provide very volatile series, which happens when the percentage of "neutral" answers is not very large and, in addition, is quite volatile itself. Finally, it is not easy to

[5] Various modifications of this setup have been proposed in the literature; see again Nardo (2003) for a survey of the literature.

generalise the probabilistic approach to the polychotomous case, because more than one scale parameter, like δ, must be set.

8.16 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 per cent, and "It did (will) increase" when x grows by 10 per cent. 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.

8.17 The regression approach has been introduced by Anderson (1952) and popularized by Pesaran (1984, 1987). Assume that $P_t = (p_{1t}, \dots, p_{kt})$ 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

$$Y_t = a_1 p_{1t} + \cdots + a_k p_{kt} + u_t \qquad (5)$$

where u_t is a random disturbance, provides a statistical estimate of the parameters a_j[6] . 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, $\hat{a} = (\hat{a}_1, \dots, \hat{a}_k)$ is given by the formula $\hat{a} = \Sigma - 1\Phi$, where Σ and Φ are the variance matrix of the time series P_t and the

[6] The constant term is dropped from the regression model, since $p_{1t} + \cdots + p_{kt} = 1$ by definition.

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 $\hat{a}$ cannot be interpreted as the average level of x attached by the interviewed persons to each qualitative answer.

8.18 The regression approach is very simple and is based on a smaller number 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 sectoral 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[7]. For further details about the methods described, refer to specific literature mentioned above in the text.

4. The quantification of qualitative information

8.19 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 and to take into account 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.

8.20 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 have allowed an analysis of survey data that may well complement the more usual one based on the calculation of the balances. In the questionnaire of the manufacturing survey monthly carried out by ISTAT, for example, the method uses individual responses to a selected number of qualitative questions.

8.21 Based on six questions selected for the analysis (the six questions consist of the following: assessments of the total, internal, and foreign orders and production; and expectations on production and orders for three months ahead), ISTAT chose to concentrate on the most economically significant behavioural models possibly emerging from the data (see the following table)[8]: this choice allows an easier interpretation of the results, allowing also the comparison of

---

[7] For empirical evidence on the performance of the methods described above see D'Elia, E.(2005).

[8] The number of behavioural models is 729. In fact, the number of all the possible permutations of the three elements in 6 questions is 3^6. However, as a logical consequence of the results obtained from the correlation

models over time. The nine models are represented by the nine centres (or representative objects) imposed in the clustering procedure[9]. In Table 8.1, the nine centres are described.

Table 8.1: Behavioural models

		Expectations		
		Positive	Unchanged	Negative
Assessments	Positive	PP	PU	PN
	Unchanged	UP	UU	UN
	Negative	NP	NU	NN

8.22 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 the expectation 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 does prevail, firms are expressing negative assessments and expectations on the six variables on which the analysis is performed. All the other models represent intermediate situations (model UP represents firms that answered "Unchanged" for all the assessments and "Positive" for all the expectations).

analysis between selected variables, ISTAT considers only the major nine common association paths shown in Table 8.1.

[9] 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.

8.23 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.

8.24 Other alternative methods of quantification were proposed 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 proposed. The first is a nonparametric method based on the notion of the spectral envelope; it originates a signal extraction filter which has a solely cross-sectional dimension, i.e. only contemporaneous values are employed. The second method 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. Further details and application of these methods can be found in Proietti and Frale (2010).

5. The "disaggregate" approach

8.25 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.

8.26 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 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.

8.27 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).

6. About the choice of the quantification method

8.28 Ideally, the choice of a quantification method should be made after an exhaustive comparison of the main different quantification methods to find out the transformation that produces a measurement error with the lowest possible size. In practice, the use of results (presentation of data, econometric analyses etc.) and their feasibility (easy-to compute and easy-to understand) should be taken into account. A compromise among all these aspects should be made. More specifically, diffusion indexes or balances may be used for the presentation of

survey results (e.g. in monthly press releases, in database etc.), for calculating composite indicators (e.g. balances are used in the confidence climate, see chapter 10), for business cycle analyses and short-term forecasting; the other approaches may be used for business cycle analyses and economic research.

B. Seasonal adjustment of tendency survey data

8.29 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 of economic, labour market and financial developments from one period to the next. There is thus a need to remove these types of influences to better understand whether opinions of producers and consumers improve, stagger or deteriorate in the short run, i.e. from month to month or quarter to quarter, and to properly and more easily detect major turning points in the business cycle (once the seasonal variations are removed).

8.30 In general, seasonal patterns might vary depending on the kind of economic activity and/or the type of indicator. Business managers are usually aware of the seasonal fluctuations in their economic activity and, even though they are asked not to take into account seasonal effects in their response, in practice respondents may be unable to do so completely or sufficiently and the answers may still show seasonal patterns.

8.31 Many outcomes of ETS variables exhibit seasonal patterns. Even though respondents are explicitly asked not to take into account such seasonal variations, in practice the answers frequently show seasonal patterns. For example, an analysis of DG ECFIN on the ETS series in the harmonised programme indicated that approximately 45 per cent of ETS series show a

seasonal pattern, 40 per cent did not show a seasonal component and around 15 per cent showed irregular behaviour (EC, 2006 p.60). The seasonal patterns in the data are usually weak and unlikely to change much over time.

8.32 In CTS, consumers are also asked not 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 wavers throughout the year.

8.33 *How to correct for seasonality*? For time series of quantitative monthly or quarterly statistics, the repeatedly seasonal patterns are adjusted by computing changes over one year where each month is compared with the corresponding month of the year before. This is one often used method to assess data showing seasonal patterns. Under the assumption that seasonal patterns are very similar or the same in each year, this method gives a simple indication of the development of the indicator over time.

8.34 Although the year-on-year method is effective to avoid the impact of seasonality, it has some serious drawbacks for the analysis of developments in the short run. While on a yearly basis the indicators might still improve or deteriorate, the month-on-month or quarter-on-quarter development may exhibit a different pattern. For instance, in almost all cases the improvement or the deterioration starts earlier or has already developed in the opposite direction. Only after explicitly removing the seasonality can developments be traced more accurately on a sub-annual basis. Furthermore, year-on-year comparisons are influenced by working and trading day effects

and possible outliers. So, to adjust for these factors, advanced adjustments with sound methods should be used to remove seasonality.

8.35 Another important drawback of year-on-year comparison is that the figures are less reliable to determine turning points in the business cycle. A peak or trough in the year-on-year growth rates simply indicates that a period of relative high or low growth has ended. Thus this method is generally not recommended to for economic or cyclical analysis.

8.36 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 data are disseminated. 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. An expert time series analyst may be required for this.

8.37 As mentioned before, it is easier to analyse changes of seasonally adjusted time series in the short run, because the seasonal swings, which can hinder the analyses, are smoothed or filtered out of the raw data. The seasonal adjustment may be performed in a *direct* or *indirect* approach. The indirect seasonal adjustment is when the seasonally adjusted estimate of a time series is derived by combining the estimates of two or more directly adjusted time series. Suppose, for example, that total manufacturing production is calculated as the sum of five underlying branches. Seasonal adjustment is applied first to each of the five underlying individual branches, and then added together to obtain the series for the seasonally adjusted total manufacturing. The direct seasonal adjustment is performed if all time series, including aggregates, are seasonally adjusted on an individual basis. For an informed choice between using the direct or indirect approach users should consider descriptive statistics on the quality of

the indirectly and directly seasonally adjusted estimates, such as, the smoothness of the component time series, residual seasonality tests on the indirect seasonally adjusted estimates, and measures of revision; the characteristics of the seasonal pattern in the component time series; and the user demand for consistent and coherent outputs, especially where they are additively related (Eurostat, 2009).

1. General principles of seasonal adjustment of BCS data

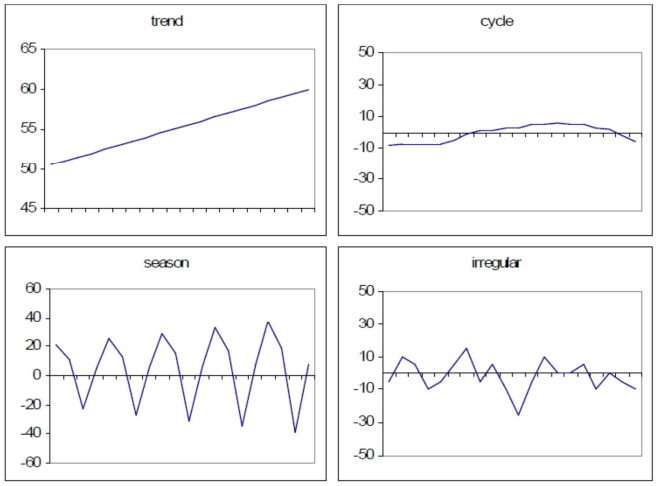
8.38 In general, series that do not show seasonality should not be adjusted. The seasonal adjustment should be guided by the following general principles:

- 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 preferably be simpler (ARIMA) models;
- The underlying choices should be documented.

8.39 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, a time series is decomposed into its separate components, so that the seasonal component can be filtered out. However, the separate components are not observable, because only the series in its totality can be observed. Seasonal adjustment is therefore subjective to a certain extent; there is more than one way of performing this decomposition.

8.40 It is assumed that a calendar and working day-adjusted time series (Y) can be broken down into the following four components: the trend component (T), the business cycle or cycle component (C), the seasonal component (S) and the irregular component (I) as shown in Figure 8.1.

Figure 8.1: Four main components in time series



Source: Statistics Netherlands, 2011, Method series 11, Seasonal adjustment

8.41 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 width of, e.g. +100 and -100 (except for a few variables such as utilisation degree or index order position). However, it may be worth emphasizing that sometimes, in shorter time series, a trend can be identified and diagnosed with the use of the most common statistical methods. The trading day and holiday component is also not used in confidence surveys: considering the qualitative nature of the data, the impact of these factors is not as significant as for the quantitative data. In ETS, the outliers, usually extracted prior to a seasonal component, are more recurring for quantitative

questions; in qualitative questions are limited by the boundaries in the answering categories (e.g. +/- 100 per cent).

8.42 The following components may be present in business and consumer tendency surveys:

a) The *seasonal* component
b) The *business cycle*
c) 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.

2. Widely used solutions for seasonal adjustment

8.43 This Section presents examples of methods commonly used for the extraction of the seasonal component. For further details and more software, we refer to the bibliography.

8.44 X12-RegARIMA, developed by the US Census, 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.

8.45 TRAMO-SEATS, developed by the Bank of Spain, is a parametric approach and software package. SEATS is the module for the seasonal adjustment. The method tries to model time series and estimate seasonal effects using ARIMA models. TRAMO makes use of regression for the same reasons as X12, trading days 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.

8.46 Another method for adjusting for seasonality is Structural Time Series Models (STM). This approach is intended to extract the seasonal adjustment simultaneously with other components such as a 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.

8.47 Dainties is currently used by DG ECFIN for adjusting 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 Annex 1 of the ECFIN'S User Guide (EC DG ECFIN, 2014).

8.48 National statistics offices increasingly make use of either X12-RegARIMA or TRAMO-SEATS where 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 seasonal adjustment, for example (Eurostat, 2009) do not prescribe a preference for either of the methods.

8.49 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 particularly suitable for tendency surveys as they are not generally revised. A drawback is that no outlier correction is currently available for the Dainties method. A 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.

8.50 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.

Chapter 9: Metadata and data dissemination

A. Introduction

9.1 Once data are collected, processed and validated, they are disseminated to the public. An important element in data dissemination is the preparation of metadata which would help users to better understand the figures and the methods used to compile the data in order to appropriately use them. Section B of this chapter, in particular, describes in more detail the type of information that is generally provided in the metadata which should accompany the statistical output (tabulations, publications, etc.) of economic tendency surveys. Section C describes data dissemination practices for ETS.

B. Metadata

9.2 Data dissemination consists not only of the dissemination of the statistical data but also the provision of metadata describing the statistical output. By providing a description of the statistical output, the quality of the output is greatly increased for users.

9.3 Metadata provide information and explanations on how, by whom and for what purpose the data are produced or generated, details on the survey process and methodology, as well as possible limitations of the data. Metadata are very useful both for data producers (to document their procedures) and data users (to better understand the data). The information provided in the metadata enables users to draw accurate inferences from the data and to determine whether the results of surveys conducted in different countries are comparable. The metadata should be updated regularly or at least as soon as there are changes in the methodology.

9.4 There are different levels of metadata information: detailed information may be useful when it is directly linked to the data (e.g. to a downloadable database) to explain the underlying concepts and definitions. These types of metadata are generally comprehensive (in terms of

description of the production process and its features) and presented in a long form. Sometimes, summary metadata information with basic information on ETS may be useful for expert users that need to check main aspects of the surveys (e.g. need to compare data of different countries, etc.).

9.5 Table 9.1 shows an example of metadata for ETS which is used in tendency surveys and contains the main aspects that should be included in summary metadata[1].

Table 9.1: ETS: an example of metadata information

Contact data
Country Survey Organization Official address Website Contact person
Methodology

[1] Most items of the table are in the metadata set supporting business and consumer surveys in the Joint EU Harmonized Programme of the EC DG ECFIN (for further details about metadata set requested by the EC see http://ec.europa.eu/economy_finance/db_indicators/surveys/metadata/index_en.htm).

Survey frame Size of frame list Characteristics Frame list update Population Sampling method Sample size Sampling error Response rate Treatment of non-responses Sample coverage Weighting scheme Kinds of economic activities (classification and categories covered) Periodicity Survey method Fieldwork period Remarks methodology
Data and dissemination
Type of data Unit of measure Data revision Frequency of dissemination

Dissemination format Release policy
Metadata update

9.6 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.

1. Contact data

9.7 The contact data represent useful information on the institution conducting the survey as well as on the contact person. In this section, the name of the institution that carries out the survey, its address and 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.

2. The methodology

9.8 This part of the metadata provides information concerning the methodology. More specifically, information about the target population, the sampling frame, sample size, accuracy etc..

9.9 The item *survey frame* (see chapter 4 for definition) provides a description of the sampling frame used for the survey and the type of units contained in the frame. For example, for BTS, this can be: the business register, government register, electoral rolls, own sources of information etc. which contain for example a list of enterprises or establishments etc. For CTS, the frame can be: list from official census, statistical registers, mobile phone register or fixed telephone lists, etc. which contain a list of households or consumers.

9.10 The item *size of frame list* refers to number of statistical units of the frame list.

9.11 The *characteristics* that are considered in the sampling frame list should be indicated: addresses and contact details of enterprises, kind of economic activity (ISIC code), number of employees, turnover etc. are typical examples for BTS; while for CTS, common examples of characteristics are: name, telephone number, age, gender, socio-economic group etc..

9.12 The item *frame list update* indicates how often the frame list is updated: every year, every two years, 3-4 years etc.

9.13 The item *population* refers to the description of the target population.

9.14 In the part devoted to sampling and accuracy, the selection method of units has to be pointed out under the item *sampling method*: probabilistic selection (representative sample, balanced sampling, random route sampling, systematic sample etc.), or deterministic one (purposive or judgemental methods, quota sample etc.). It is important to describe sample design features: e.g., the presence of stages or stratification needs 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 the sampling methods see chapter 4). *Sample size* refers to the predefined total number of responses to obtain in each survey round. *Sampling error* refers to the desired precision of the estimates (for more details see chapter 5). 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.

9.15 The item *response rate* refers to the response rate of the survey. For BTS, it can be useful to 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).

9.16 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).

9.17 The *sample coverage* means the extent or area covered by the sampling method. It may be calculated in terms of units (expressions such as "50 per cent 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.

9.18 Under item *weighting scheme*, the weights used in the process of integration from the sample to universe have to be made explicit (for more details on the choice of weights and process of estimation for business and consumer tendency surveys see chapter 5).

9.19 In the item *kinds of economic activities (classification and categories covered)*, the classification of economic activities used for BTS should be made explicit (for example, ISIC rev.4) as well as or categories (for BTS, division, group of ISIC; for CTS, income classes, age groups etc.) covered by the survey and the available industry breakdowns have to be stated.

9.20 As for the part regarding data collection, under item *periodicity*, the frequency of the survey has to be pointed out. It should be noted that some questions may have different frequencies with respect to the survey (e.g. the survey has a monthly frequency but some questions are asked quarterly or biannually): the periodicity and the time during which these questions are asked should be indicated clearly in the metadata.

9.21 *Survey method* describes the method used for data collection: CAPI, CATI, CASI etc. For the most common methods used for business and consumer tendency surveys see chapter 6.

9.22 *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.).

9.23 *Remarks methodology*. This item may be used to indicate any useful remark related to the methodology, such as, for example, "Quarterly questions are asked during the January, April, July and October survey" or some information concerning the quality of the list used for sampling scheme[2].

3. Data and dissemination

9.24 In *data and dissemination*, additional information about data, dissemination of results and release policy are requested. In *type of data*, percentage of reply options, balances, composite indicators (e.g. confidence climates) may be indicated.

9.25 The *unit of measure* of each indicator has 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. Balances can also be indicated in this item (for the calculation of balances see chapter 8).

9.26 *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.

---

[2] For example the remark in CTS for France (INSEE) is the following one: "The trading list of Pages Jaunes Marketing Service is not completely exhaustive: a number of households ask Pages Jaunes Marketing Service not to appear in the public directory; the Pages Jaunes Marketing Service list does not include cell phone numbers; neither does it cover the population of clients of other phone operators. The survey sampling plan includes a correction procedure taking account of the incomplete coverage of the frame list, which enables INSEE to extrapolate results relating to the whole population of households".

9.27 The *frequency of dissemination* should be indicated: monthly, quarterly etc.

9.28 *Dissemination format*: all the information concerning formats used for the data dissemination has to be described: publications, online database, micro data access, or other (e.g. website address etc.).

9.29 In the item *Release policy*, information about the release calendar (specify when the calendar is updated) and users' access to the calendar may be described.

9.30 *Metadata update*: the date metadata were last certified, last posted and last updated should be stated.

C. Data dissemination

1. Variety of users

9.31 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.

2. Quality criteria

9.32 Following dissemination procedures with a 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 standardization 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).

3. Release calendar

9.33 Releasing publications and results according to a yearly calendar that is disseminated well in advance is an obvious way to increase the use of tendency surveys. It also helps to ensure punctuality 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. For example, in the Joint Harmonised EU Programme, 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. Standardized publications are being released in the same way in every reporting period, which is convenient for users as well as statistical offices.

4. Forms of dissemination

9.34 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 should always be available to increase the transparency and usefulness of the data.

9.35 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 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.

9.36 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 Joint Harmonised EU Programme, for example, two separate press releases are sent out to news agencies for the Economic Sentiment Indicator (ESI) and the Business Climate Indicator (BCI).

9.37 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 the recent past and some other statistical indicators, such as long-term averages and min/max values. In the Joint Harmonised EU Programme, for example, 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.

9.38 Special reports can be provided on demand to respondents to the survey or other parties if allowed by the statistical law. Special reports with detailed results on the specific industry of individual respondents provide an important incentive for managers to respond to the survey.

Interested respondents can benchmark their own results within their industry. 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.

9.39 *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 industry or regional breakdowns, others are even interested in specific questions, such as 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 on the web simultaneously with the press release in the form of tables containing long time series on all composite indicators (e.g. confidence climates), underlying balances, reply options (in percentage) to each question, including, if available, breakdowns for industry divisions/groups or consumer categories. Clear descriptions (e.g., seasonally adjusted or raw data, etc.) and extra explanation of the data should accompany the tables and database.

9.40 The development and maintenance of a database for online data sharing requires specialised software. Different types of users can access information suited 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. Websites offer the possibility of more interactive and dynamic (animated) applications to present the results of the surveys. Balances can be sometimes difficult to interpret and compare across

kinds of economic activities and regions, even for regular users of the data. Therefore, a link to an exhaustive metadata file (in a long form), explaining the underlying concepts and definitions is vital. This metadata file should also describe the production process and its features: information content, survey phases and operations, activities to prevent, monitor and evaluate sampling and non-sampling errors. Statistical information systems for data dissemination and their sources should be also documented[3].

9.41 Finally, a reference should be made to the so called Forward Looking Indicators (FLIs) used in the IMF's Special Data Dissemination Standard (SDDS). According to the SDDS guide, subscribers are encouraged to disseminate more structured and focused FLIs that include, among others, surveys of expectations, such as BTS and CTS. The SDDS enhances the dissemination of FLIs because they can provide useful insights into the developments of an economy; these indicators have been already introduced by some EU countries.

5. Confidentiality

9.42 Individual data gathered with business and consumer tendency surveys are generally 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 industry data are published, there is a risk of exposing information of individual businesses.

9.43 Both primary and secondary confidentiality should be guarded to avoid unwanted revelations. Primary confidentiality is data of a unit which can be directly identified. Secondary confidentiality is information on a unit that can be identified when combined with other data. If

[3] In Italy, this type of metadata information is available on the website. For further details see http://siqual.istat.it/SIQual/lang.do?language=UK

micro-data is made available for (academic) research purposes, it should always be made anonymous.

6. Revisions

9.44 Generally, there are no revisions in ETS data due to extra information from incoming data [4]. However, revisions may take place due to change in methodology such as updating of the weights for regional aggregates. Revisions can also take place for other reasons, planned or unplanned. For instance, changes in the method used for seasonal adjustment (see Chapter 8) or changes in the classifications of economic activities can also lead to revisions. As in all statistics, mistakes in published data may also require revisions.

[4] In the joint EU programme, an exception is, for example, the French survey programme ran by the INSEE, where late replies (received after the month's cut-off date) relating to month *m* are included in the *m* month's results at the end of the month *m+1*, thereby leading to (slight) revisions.

Chapter 10: Use of tendency survey results

A. Users of tendency survey data

10.1 For years, the main objective of ETS has been to collect information on business conditions for the benefit of respondents. These surveys were therefore carried out mainly by trade associations. Nowadays, ETS are considered 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 governmental bodies are involved in the collection and analysis of these data. ETS results are also closely monitored by financial analysts and often cited in newspaper articles.

10.2 Four categories of potential users of tendency surveys can be broadly identified as follows (see also OECD, 2003):

(a) international organizations 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, and financial institutions;

(d) the press and other media.

10.3 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 academia, research institutions and governmental bodies. The second category may be labelled "executives" and consists of senior business executives, politicians, senior civil servants responsible for government policy, senior personnel in banks and financial institutions and the press and other media. "Executives" probably make up the majority of BTS data users.

10.4 At the international level the EC, for example, could 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.

10.5 Other important users among the international governmental organizations are the OECD and the ECB. The OECD derives for almost every member state an indicator also based on survey data with good leading properties. It also selects headline indicators (for industry and consumers) from the business and consumer tendency survey database (and from other sources for non-EU countries) and combines them to create the Composite Leading Indicators. OECD goes into question level detail, such as for order books and employment, but does not use survey data for sub-sectors. It also regularly publishes survey results in conjunction with "hard data" in the Main Economic Indicators database. Also the ECB regularly uses 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.

10.6 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[1] and the University of Michigan[2] publish and analyse indicators referred to the US economy, while the Bank of Japan[3] conducts the Tankan survey for the Japanese corporate sector.

10.7 "Executive" users also 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 interested above all 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 use survey products in quantitative analyses such as in forecasting and macroeconomic modelling. In some cases, particularly among financial institutions, ETS data are also used to inform business/investment decisions. Finally, last but not

[1] https://www.conference-board.org/
[2] http://www.sca.isr.umich.edu/
[3] http://www.boj.or.jp/en/statistics/tk/index.htm/

least, the biennial CIRET conferences may be considered as a good opportunity for a broad worldwide dissemination of survey data.

B. Uses of tendency survey data

10.8 Tendency surveys are implicitly based on the assumption that insiders are in a privileged position to assess developments in a given market. Therefore, if taken on an aggregate basis, the insight of the insiders can roughly anticipate business cycle trends. The surveys came into vogue as a purely empirical instrument to track the business cycle with the 1929 recession and they 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 is still nowadays a matter of fierce debate among economic scholars. Nevertheless, there is quite a copious volume of well documented studies that recognize tendency survey data as a crucial instrument for gathering information in today's ever-changing environment (see among the others, Koopmans, 1947; Zarnowitz, 1992).

10.9 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 tools to visualize 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' behaviour.

10.10 The various uses of ETS data are mainly based on the fact that the statistical series derived from such surveys are by their nature particularly suitable for business cycle monitoring

and forecasting. There is substantial evidence that survey series are a good proxy for corresponding quantitative series, if available, and show good relationships with some general reference series representing the business cycle or general economic development, like industrial production or GDP.

10.11 The relationship between quantitative and ETS data was studied early in the history of ETS and it was clear that the main problem was to find out a criterion to compare ETS data (balance series) with quantitative series. The comparison between survey and quantitative data is made difficult by the fact that quantitative series are expressed in value or volume terms, while ETS use ordinal scales. In this context, the main issue is to focus the concept of business cycle explained by the balance series. 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 refer to the concept of classical, growth or growth rate cycle. The first defines a recession as a decline in the absolute level of a series, the second as a decline in the de-trended level and the third as a decline in the growth rate cycle series. ETS data are typically considered in the growth cycle or growth rate cycle framework because of their trend-free nature. Generally speaking, questions asking whether present levels of activity are above normal, normal or below normal have a balance that represent a growth cycle. The related quantitative series may be de-trended[4] for a correct comparison. For questions asking for an assessment on present or future changes or trend in comparison to past or present periods, the balance series corresponds to growth rate cycle of the quantitative data. In this case, the difficulty is that respondents may not actually use the reference

[4] For de-trending methods see the forthcoming UN Handbook on Cyclical Composite Indicators. The finalization is expected by 2015.

period specified in the questionnaire in answering the questions. When a quantitative series is converted to monthly or quarterly changes it may not correspond with the reference period used by respondents answering questions about present or future changes (often annual percentage growth rates are necessary for a quantitative series).

10.12 The analyses indicate that balance series reflect changes over longer periods than the one requested in the survey (OECD, 2003). Practically, after an appropriate transformation of quantitative series (de-trended or annual percentage growth rates), a first graphic comparison between balances and quantitative series may be made. Furthermore, the relationship may be enhanced analyzing the performance of ETS series in tracking their reference quantitative series (e.g. to test their cyclical features in order to include them in composite indicators) as described in the following section.

1. Business cycle composite indicators

10.13 Composite indicators are formed when individual indicators are compiled into a single index on the basis of an underlying model of the multidimensional concept that is being measured (OECD – glossary of statistical terms)[5]. 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 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. Composite indicators may be leading, coincident and lagging indicators of

[5] For further details see also OECD (2008).

economic activity (this classification has a long-lived tradition in economic research – Burns and Mitchell, 1946).

10.14 Generally speaking, there are various approaches to build composite indicators. For a comprehensive review of the methods, we refer to the UN Handbook on Cyclical Composite Indicators. Here only three commonly used methods are mentioned: 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 objective of the indicators.

10.15 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 widely used in Europe and in the US, variables are selected relying on descriptive statistics and on the detection of turning points. More specifically, balance series may be evaluated for time consistency, conformity and economic significance[6]. The time consistency is evaluated by calculating the average lead/lag of the survey series at turning points, identified with the Bry-Boschan routine. For conformity it is possible to calculate an indicator of directional coherence consisting in the percentage of cases where the balances show the same movement (plus or minus) as that of the quantitative series. Finally, cross-correlation function with respect to the reference series may be calculated to assess the economic significance. Once relevant variables have been selected, a simple way to aggregate them is to compute the arithmetic mean of net balances, as is the case for the computation of confidence indicators. The general idea behind the construction of such indicators is that each survey answer contains a common component which

[6] For further details about the criteria of selecting series see chapter 6 of the forthcoming UN Handbook on Cyclical Composite Indicators.

can be better extracted by a cross-sectional average. A weighting scheme may also be chosen to aggregate between different sectors and, if needed, between different countries.

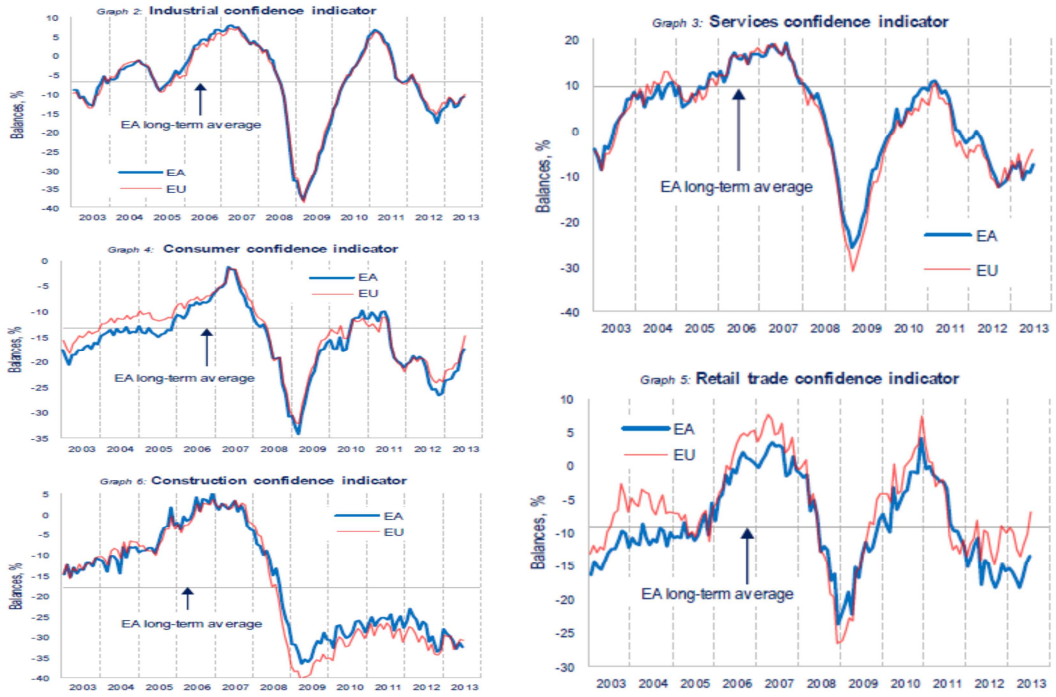
10.16 Within this approach, survey producers in EU countries, for example, have designed confidence indicators maximizing the coincident correlation with the reference series conveniently changed (for instance the year-on-year growth rate of industrial production in the case of the industry survey). Hence, the confidence indicators computed in EU countries can be considered as composite coincident indicators for each sector.

10.17 More specifically for each kind of economic activities surveyed, five separate confidence indicators are elaborated regarding the following variables (EC DG ECFIN, 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, unemployment expectations (with the inverted sign) and savings, over the next 12 months in each case.

10.18 The EC then calculates the EU and euro-area aggregates on the basis of national results and seasonally adjusts the balance series (see Figure 10.1). 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 constructing such indicators is justified because it may improve evaluation of business cycle conditions -for example to monitor kinds of economic activities that typically anticipate the overall cycle, such as construction, or that represent important engines of growth, such as services or consumption from a demand perspective.

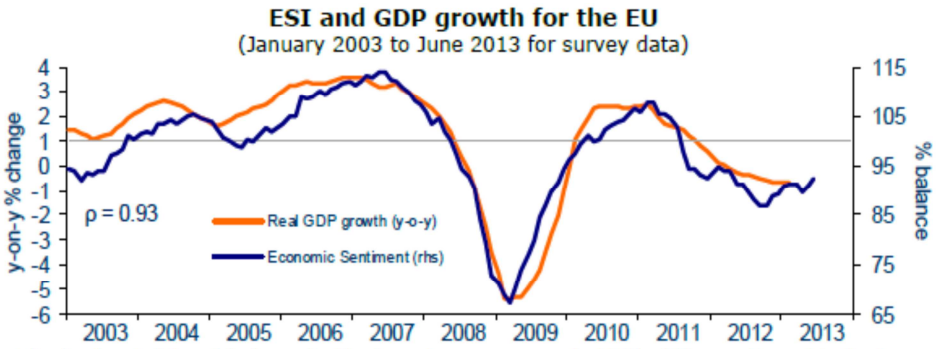
Figure 10.1: Confidence indicators



Source: European Commission, Business and Consumer surveys

10.19 Accordingly, alongside these indicators, the EC also provides the so-called Economic Sentiment Indicator (ESI) as a weighted average of the balances of replies to some selected questions addressed to firms and consumers in the five areas 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 areas 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 10.2).

Figure 10.2: Economic Sentiment Indicator



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.

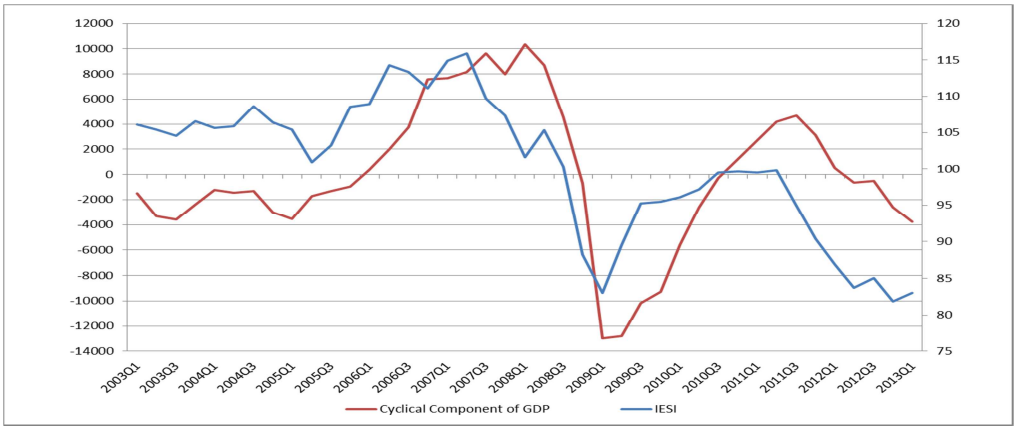
Source: European Business Cycle indicators

10.20 Furthermore, alongside the business cycle indicators, a special and prominent importance is devoted to leading indicators, mainly because of 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 behaviour 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).

10.21 The relevance of such indexes encourages almost all 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, wholesale and retail trade industries). 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 is 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 10.3).

Figure 10.3: IESI and cyclical component (Hodrick-Prescott) of GDP

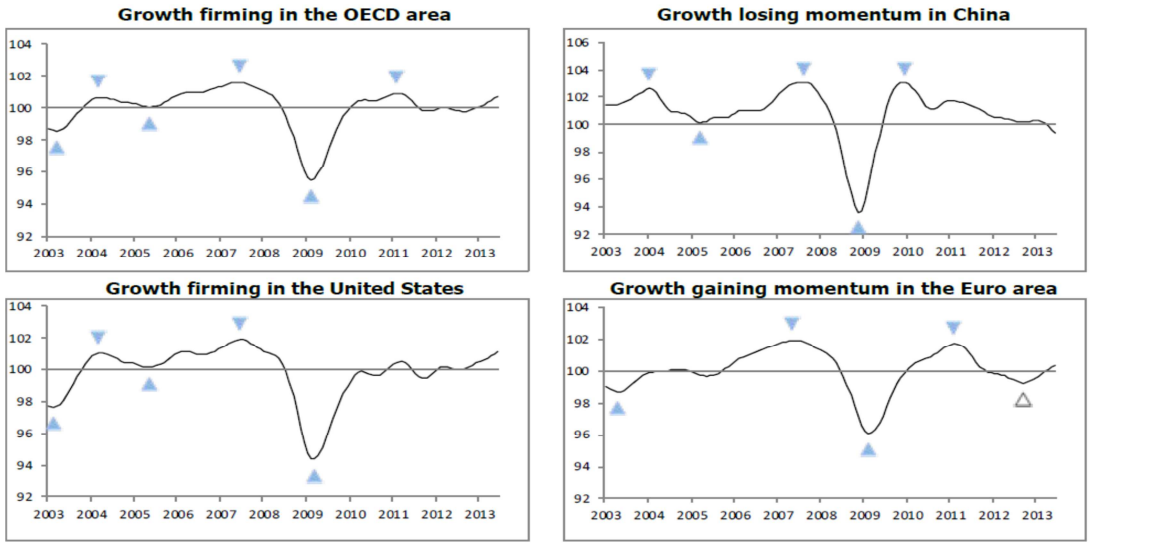


Source: ISTAT

10.22 Finally, BTS data can be also combined with quantitative statistics to obtain more structured cyclical indicators. For example, the OECD system of Composite Leading Indicators (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 10.4).

Figure 10.4: 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.

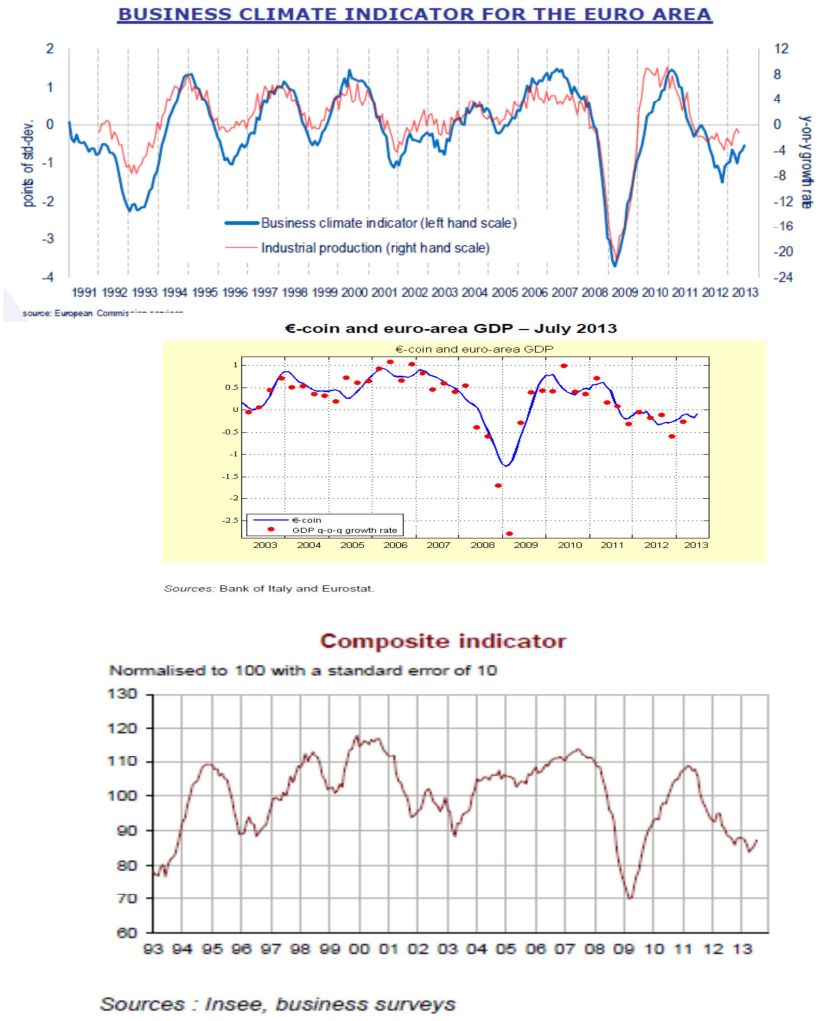
Source: OECD Composite Leading Indicators News Release

10.23 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 of 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 behaviour of the variables during expansions and recessions.

10.24 Widely used empirical methodologies are developed along the lines of 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 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 survey indicators (see on this issue the examples displayed in Figure 10.5).

Figure 10.5: Examples of factor-based composite coincident indicators



10.25 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 industry and limited gains for the other kinds of economic activities. However, alongside more specific studies about dating cycle analysis, Hamilton's (1990) MSM provides a more convenient statistical framework. Basically, the

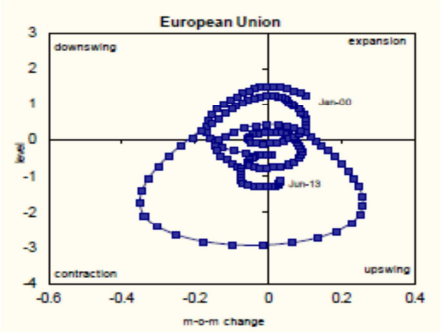
composite indicator in this context represents the status of the business cycle (expansion/recession mapped into a binary variable), which determines the behaviour of all the coincident indicators that can change substantially over different phases of the cycle. Diebold and Rudebusch (1996) suggest that these two approaches can be combined by allowing the underlying factor in the Stock and Watson model to evolve according to a MSM.

2. Visualizing tools

10.26 Recently, business cycle indicators are also used to convey an immediate, self-explanatory visual representation of the business cycle situation, through the construction of so-called visualizing tools. The EC produces the "Economic Climate Tracer" (Gayer, 2010) aimed at visualizing the state of the economy using industry 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'[7]. 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 graph. The latter directly corresponds to a turning-point oriented definition of the business cycle phases (Figure 10.6).

[7] http://ec.europa.eu/economy_finance/publications/cycle_indicators/

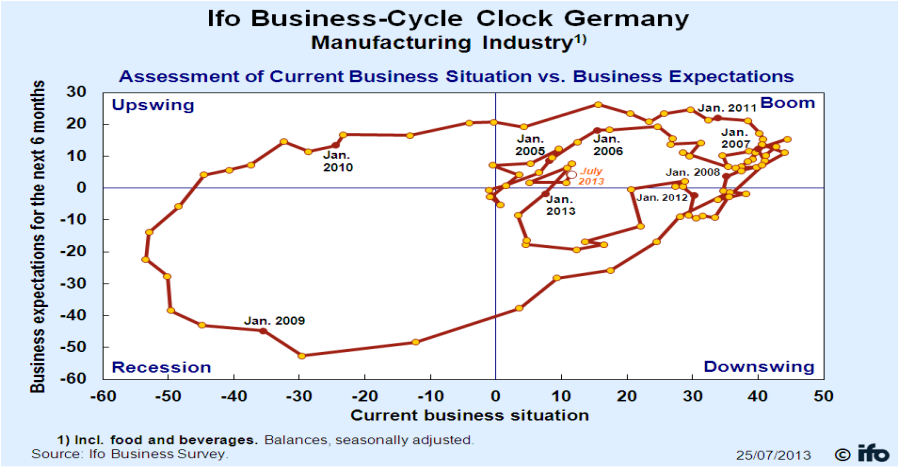
Figure 10.6: Economic Tracer



Source: European Business Cycle indicators

10.27 Various other graphical monitoring systems that display the cyclical component of selected indicators and their modification (vis-à-vis the previous period) in a four 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 10.7) 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 10.7: IFO Business Cycle Clock



3. Forecasting using survey data

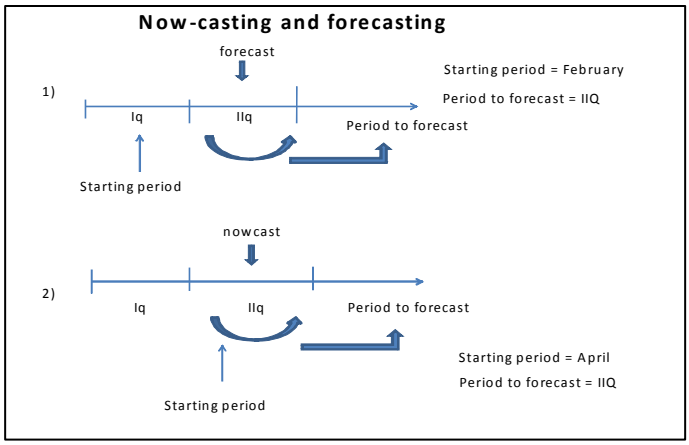
10.28 From a general point of view, the rationale for selecting indicators to be included in forecasting models is essentially based on their statistical characteristics. In this sense, timeliness and absence of revision, immediate reliability and a significant correlation with the economic variable 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, consumption expenditure or industrial production.

10.29 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 $t+1$. Therefore, even if the survey indicator is strictly coincident, its use allows a one-step-ahead forecast of the industrial production index because of the availability of the indicator.

10.30 A number of econometric tools have recently been developed to predict economic behaviour in the current and immediately following quarters. These kinds of exercises are commonly referred to as nowcasting and short-term forecasting respectively, (see Figure 10.8 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 10.8: Nowcasting and forecasting



10.31 For example, as far as indicator selection is concerned, the activity of extracting reliable signals from high frequency indicators can be carried out according to two main directions: the *factor models* and the *bridge models*. As said before, factor models 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 bridge models 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 bridge models require that the whole set of regressors should be known over the projection period, these kinds 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 broad usage, both the factor model and bridge model approaches are subject to criticism. Factor models may be biased by

unbalanced sources of information (Boivin and Ng, 2006), while bridge models 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 bridge models may often be quite difficult to replicate.

10.32 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 predict consumer spending (Crushore, 2005).

4. Consumer confidence, macroeconomic analysis and business cycle

10.33 The role of confidence indicators in economic analysis has been widely debated in recent 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. This section provides a brief overview of the existing literature; this is intended as an introduction to the debate and a starting point for further research, without any ambition to exhaustiveness.

10.34 Broadly speaking, the link between confidence and economic decisions has usually been 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).

10.35 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, such 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).

10.36 Empirical results on effective forecasting ability of consumer confidence are, however, mixed and still inconclusive (see among 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).

10.37 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 for modelling the relationship of consumer attitudes to confidence. Using 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 research has recently proposed using 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 which avoids too restrictive assumptions about functional form in order to explore the usefulness of confidence to forecast consumption. Focusing on Italian data, this relationship is characterized by asymmetries over the business cycle. The empirical results show that expected consumption growth is always very negative when the confidence falls below certain threshold, while the opposite 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.

10.38 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 signalling 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 wars; see on this issue Throop, 1992 and Garner, 2002, respectively). Acemoglu and Scott (1994) suggest that the

circumstance that UK consumer confidence predicts consumption even after controlling on income may be due to shifts of the consumption function over the business cycle.

10.39 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 impulses behind fluctuations. In their paper, they find that these kinds 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 modelling 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).

10.40 Some other authors also underline that 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 (commensurately reducing 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.

10.41 Recently, confidence indicators have also been used in literature concerning international synchronization of business cycles, both at the global level (Kose et al., 2012) and the regional

level (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 at the global and local level. One of the main reasons 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 viewed 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.

10.42 Finally, substantial discrepancies in activity at industry level justify the interest - along the business cycle analysis - in an industry disaggregation. In this context, ETS data may also be 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 and availability of industry information. In particular, comparing the performance of tradable and non-tradable types of economic activities 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 industrial shifts are taking place in national economies, the adjustment moving in the expected direction according to standard economic theory.

5. Micro-econometric methods

10.43 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 the 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 make it possible 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 consumer 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.

10.44 Many research 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 are the professional forecasts 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) using ISTAT CTS data found 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.

10.45 Departures from the Muthian rationality hypothesis have also been investigated using survey data by Bovi (2009 and 2012). In his first contribution, Bovi used a large dataset covering ten European countries over 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 able to enhance the forecasting ability of even the more statistically efficient econometric models, providing further contradiction to the standard hypothesis of rationality.

10.46 More recently, various authors have also used micro-level information stemming from BTS. Rottman and Wolmershauer (2013) used data stemming from the monthly IFO Business survey on the manufacturing industry in order to derive an aggregate indicator of a credit crunch

in 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 over 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 were also able to derive a credit crunch indicator, which measures the part of the shift in the loan supply that is explained neither by firm-specific factors nor by the opportunity costs of providing risky loans. Similar results for Italy were found by Costa et al. (2012) using the ISTAT BTS data. They were also able to extend the analysis to the services, retail and construction industries thus using a richer set of controls for characterizing firms' creditworthiness.

10.47 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 industry 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 focusing 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) found 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.

Annex 1: Detailed description of economic activities covered in BTS

The table below shows the correspondence of the economic activities covered in BTS and the International Standard Industrial Classification of All Economic Activities, (ISIC Rev.4). In particular, it shows the level of detail (at the 2-digit level) of economic activities covered in each group.

Core Economic Activities	**ISIC Rev. 4**
Manufacturing	C - Manufacturing 10 - Manufacture of food products 11 - Manufacture of beverages 12 - Manufacture of tobacco products 13 - Manufacture of textiles 14 - Manufacture of wearing apparel 15 - Manufacture of leather and related products 16 - Manufacture of wood and of products of wood and cork, except furniture; manufacture of articles of straw and plaiting materials 17 - Manufacture of paper and paper products 18 - Printing and reproduction of recorded media 19 - Manufacture of coke and refined petroleum products 20 - Manufacture of chemicals and chemical products 21 - Manufacture of basic pharmaceutical products and pharmaceutical preparations 22 - Manufacture of rubber and plastics products 23 - Manufacture of other non-metallic mineral products 24 - Manufacture of basic metals 25 - Manufacture of fabricated metal products, except machinery and equipment 26 - Manufacture of computer, electronic and optical products 27 - Manufacture of electrical equipment 28 - Manufacture of machinery and equipment n.e.c. 29 - Manufacture of motor vehicles, trailers and semi-trailers 30 - Manufacture of other transport equipment 31 - Manufacture of furniture 32 - Other manufacturing 33 - Repair and installation of machinery and equipment
Construction	F - Construction 41 - Construction of buildings 42 - Civil engineering 43 - Specialized construction activities

Retail trade	G - Wholesale and retail trade; repair of motor vehicles and motorcycles[46] 45 - Wholesale and retail trade and repair of motor vehicles and motorcycles 47 - Retail trade, except of motor vehicles and motorcycles
Services	H - Transportation and storage 49 - Land transport and transport via pipelines 50 - Water transport 51 - Air transport 52 - Warehousing and support activities for transportation 53 - Postal and courier activities I - Accommodation and food service activities 55 - Accommodation 56 - Food and beverage service activities J - Information and communication 58 - Publishing activities 59 - Motion picture, video and television programme production, sound recording and music publishing activities 60 - Programming and broadcasting activities 61 - Telecommunications 62 - Computer programming, consultancy and related activities 63 - Information service activities K - Financial and insurance activities 64 - Financial service activities, except insurance and pension funding 65 - Insurance, reinsurance and pension funding, except compulsory social security 66 - Activities auxiliary to financial service and insurance activities L - Real estate activities 68 - Real estate activities M - Professional, scientific and technical activities 69 - Legal and accounting activities 70 - Activities of head offices; management consultancy activities 71 - Architectural and engineering activities; technical testing and analysis 72 - Scientific research and development 73 - Advertising and market research 74 - Other professional, scientific and technical activities 75 - Veterinary activities N - Administrative and support service activities 77 - Rental and leasing activities 78 - Employment activities 79 - Travel agency, tour operator, reservation service and related activities 80 - Security and investigation activities 81 - Services to buildings and landscape activities 82 - Office administrative, office support and other business support activities R - Arts, entertainment and recreation 90 - Creative, arts and entertainment activities

[46] Note that not the whole Section G of ISIC Rev. 4 is covered here as it excludes Division: 46: Wholesale trade, except of motor vehicles and motorcycles.

	91 - Libraries, archives, museums and other cultural activities 92 - Gambling and betting activities 93 - Sports activities and amusement and recreation activities S - Other service activities 94 - Activities of membership organizations 95 - Repair of computers and personal and household goods 96 - Other personal service activities
Activities of special interest	**ISIC Rev. 4**
Agriculture	A - Agriculture, forestry and fishing
Wholesale	46 - Wholesale trade, except of motor vehicles and motorcycles
Financial Services	K - Financial and insurance activities 64 - Financial service activities, except insurance and pension funding 65 - Insurance, reinsurance and pension funding, except compulsory social security 66 - Activities auxiliary to financial service and insurance activities

Annex 2: Standard questionnaires

A. Business Tendency Surveys[47]

1. Core Economic Activities

a) Manufacturing

Monthly questions	Topic of the question
Question: How has your production changed over the past 3 months? Answer: It has (1) increased, (2) remained unchanged, (3) decreased	Evolution of production - past 3 months
Question: How do you expect your production to change over the next 3 months? Answer: It will (1) increase, (2) remain unchanged, (3) decrease	Evolution of production - next 3 months
Question: How has your firm's total employment changed over the past 3 months? Answer: It has (1) increased, (2) remained unchanged, (3) decreased	Evolution of employment - past 3 months
Question: How do you expect your firm's total employment to change over the next 3 months? Answer: It will (1) increase, (2) remain unchanged, (3) decrease	Evolution of employment - next 3 months
Question: Do you consider your current overall order books to be..? Answer: (1) Above normal, (2) normal for the season, (3) below normal	Level of order books - current
Question: Do you consider your current level of export order books to be..? Answer: (1) Above normal, (2) normal for the season, (3) below normal	Level of export order books - current
Question: Do you consider your current stock of finished products to be..? Answer: (1) Above normal, (2) normal for the season, (3) below normal	Stock of finished goods - current
Question: How do you expect your selling prices to change over the next 3 months? Answer: It will (1) increase, (2) remain unchanged, (3) decrease	Evolution of selling prices - next 3 months

[47] As mentioned in Chapter 3 of this Handbook, it should be made clear in the introduction of each questionnaire that respondents should exclude usual seasonal variations when they answer. Otherwise, each question should start with “excluding seasonal variations”.

Quarterly questions	**Topic of the question**
Question: How have your orders changed over the past 3 months? Answer: They have (1) increased, (2) remained unchanged, (3) decreased	Evolution of order books - past 3 months
Question: How do you expect your order books to change over the next 3 months? Answer: They will (1) increase, (2) remain unchanged, (3) decrease	Evolution of order books - next 3 months
Question: How do you expect your export orders to change over the next 3 months? Answer: They will (1) increase, (2) remain unchanged, (3) decrease	Evolution of export order books - next 3 months
Question: How many months of production are assured by your current order books? Answer: please provide number of months	Months of production assured by current order books
Question: What main factors are currently limiting your production? Answer: (1) None, (2) insufficient demand, (3) shortage of labour force, (4) shortage of material, (5) financial constraints, (6) other factors	Factors limiting production - current
Question: Considering your current order books and the expected change in demand over the coming months, how do you assess your current production capacity?" Answer: (1) more than sufficient, (2) sufficient, (3) not sufficient	Production capacity - current
Question: At what capacity is your company currently operating (as a % of full capacity)? Answer: please provide %	Capacity utilisation - current
Question: Do you consider your current stock of raw materials to be..? Answer: (1) Above normal, (2) normal for the season, (3) below normal	Stock of raw materials - current
Question: How has your competitive position on the domestic market changed over the last 3 months? Answer: It has (1) improved, (2) remained unchanged, (3) deteriorated	Evolution of the competitive position on the domestic market - past 3 months

Semi-annual questions	**Topic of the question**
Question: Please state percentage change in investment last year (t-1) on investment 2 years ago (t-2): Answer: please provide %	Development of fixed investment - past year

Question: Please state percentage change in investment this year (t) on investment last year (t-1): Answer: please provide %	Development of fixed investment - this year
Question: Please state percentage change in investment next year (t+1) on investment this year (t): Answer: please provide %	Development of fixed investment - next year

b) Construction

Monthly questions	**Topic of the question**
Question: How has your building activity changed over the past 3 months? Answer: It has (1) increased, (2) remained unchanged, (3) decreased	Evolution of building activity - past 3 months
Question: How do you expect your building activity to change over the next 3 months? Answer: It will (1) increase, (2) remain unchanged, (3) decrease	Evolution of building activity - next 3 months
Question: How has your firm's total employment changed over the past 3 months? Answer: It has (1) increased, (2) remained unchanged, (3) decreased	Evolution of employment - past 3 months
Question: How do you expect your firm's total employment to change over the next 3 months? Answer: It will (1) increase, (2) remain unchanged, (3) decrease	Evolution of employment - next 3 months
Question: Do you consider your current overall order books to be..? Answer: (1) Above normal, (2) normal for the season, (3) below normal	Level of order books - current
Question: What main factors are currently limiting your building activity? Answer: (1) None, (2) insufficient demand, (3) weather conditions, (4) shortage of Labour Force, (5) shortage of material, (6) financial constraints, (7) other factors	Factors limiting production - current
Question: How do you expect your selling prices to change over the next 3 months? Answer: They will (1) increase, (2) remain unchanged, (3) decrease	Evolution of selling prices - next 3 months

Quarterly questions	**Topic of the question**
Question: Assuming normal working hours, about how many months of work is accounted for by the work in hand and the work already contracted for? Answer: please provide number of months	Months of production assured by current order books - current

Question: At what capacity is your company currently operating (as a % of full capacity)? Answer: please provide %	Capacity utilisation - current

Semi-annual questions	***Topic of the question***
Question: Please state percentage change in investment last year (t-1) on investment 2 years ago (t-2): Answer: please provide %	Development of fixed investment - past year
Question: Please state percentage change in investment this year (t) on investment last year (t-1): Answer: please provide %	Development of fixed investment - this year
Question: Please state percentage change in investment next year (t+1) on investment this year (t): Answer: please provide %	Development of fixed investment - next year

c) Retail Trade

Monthly questions	**Topic of the question**
Question: How has your business activity (sales) changed over the past 3 months? Answer: It has (1) increased, (2) remained unchanged, (3) decreased	Evolution of sales / business activity - past 3 months
Question: How do you expect your business activity (sales) to change over the next 3 months? Answer: It will (1) increase, (2) remain unchanged, (3) decrease	Evolution of sales / business activity - next 3 months
Question: How has your firm's total employment changed over the past 3 months? Answer: It has (1) increased, (2) remained unchanged, (3) decreased	Evolution of employment - past 3 months
Question: How do you expect your firm's total employment to change over the next 3 months? Answer: It will (1) increase, (2) remain unchanged, (3) decrease	Evolution of employment - next 3 months
Question: How do you expect your orders placed with suppliers to change over the next 3 months? Answer: They will [(1) increase, (2) remain unchanged, (3) decrease	Evolution of orders placed with suppliers - next 3 months
Question: Do you consider your current stock of finished products to be..? Answer: (1) Above normal, (2) normal for the season, (3)	Stock of finished goods - current

below normal	
Question: How do you expect your selling prices to change over the next 3 months? Answer: They will (1) increase, (2) remain unchanged, (3) decrease	Evolution of selling prices - next 3 months

d) Services

Monthly questions	**Topic of the question**
Question: How has your business situation changed over the past 3 months? Answer: It has (1) improved, (2) remained unchanged, (3) deteriorated	Evolution of business situation - past 3 months
Question: How do you expect your business situation to change over the next 3 months? Answer: It will (1) improve, (2) remain unchanged, (3) deteriorate	Evolution of business situation - next 3 months
Question: How has your firm's total employment changed over the past 3 months? Answer: It has (1) increased, (2) remained unchanged, (3) decreased	Evolution of employment - last 3 months
Question: How do you expect your firm's total employment to change over the next 3 months? Answer: It will (1) increase, (2) remain unchanged, (3) decrease	Evolution of employment - next 3 months
Question: How has demand (turnover) for your company's services changed over the past 3 months? Answer: It has (1) increased, (2) remained unchanged, (3) decreased	Evolution of demand (turnover) - past 3 months
Question: How do you expect the demand (turnover) for your company's services to change over the next 3 months? Answer: It will (1) increase, (2) remain unchanged, (3) decrease	Evolution of demand (turnover) - next 3 months
Question: How do you expect your selling prices to change over the next 3 months? Answer: They will (1) increase, (2) remain unchanged, (3) decrease	Evolution of selling prices - next 3 months

Quarterly questions	**Topic of the question**
Question: What main factors are currently limiting your activity? Answer: (1) None, (2) insufficient demand, (3) shortage of	Factors limiting activity - current

labour force, (4) shortage of space/ equipment, (5) financial constraints, (6) other factors	
Question: If the demand expanded, could you increase your volume of activity with your present resources? Answer: (1) Yes, (2) No, (3) If so, by how much? ...%).	Volume of activity - current

2. Activities of special interest

a) Agriculture

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

Semi-annual questions	**Topic of the question**

Question: Please state percentage change in investment last year (t-1) on investment 2 years ago (t-2): Answer: please provide %	Development of fixed investment - past year
Question: Please state percentage change in investment this year (t) on investment last year (t-1): Answer: please provide %	Development of fixed investment - this year
Question: Please state percentage change in investment next year (t+1) on investment this year (t): Answer: please provide %	Development of fixed investment - next year

b) Wholesale

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

Semi-annual questions	**Topic of the question**
Question: Please state percentage change in investment last year (t-1) on investment 2 years ago (t-2):	Development of fixed investment - past year

Answer: please provide %	
Question: Please state percentage change in investment this year (t) on investment last year (t-1): Answer: please provide %	Development of fixed investment - this year
Question: Please state percentage change in investment next year (t+1) on investment this year (t): Answer: please provide %	Development of fixed investment - next year

c) Financial Services

Monthly questions	**Topic of the question**
Question: How has your business situation changed over the past 3 months? Answer: It has (1) improved, (2) remained unchanged, (3) deteriorated	Evolution of business /situation - past 3 months
Question: How do you expect your business situation to change over the next 3 months? Answer: It will (1) improve, (2) remain unchanged, (3) deteriorate	Evolution of business situation - next 3 months
Question: How has your firm's total employment changed over the past 3 months? Answer: It has (1) increased, (2) remained unchanged, (3) decreased	Evolution of employment - past 3 months
Question: How do you expect your firm's total employment to change over the next 3 months? Answer: It will (1) increase, (2) remain unchanged, (3) decrease	Evolution of employment - next 3 months
Question: How has demand (turnover) for your company's services changed over the past 3 months? Answer: It has (1) increased, (2) remained unchanged, (3) decreased	Evolution of demand (turnover) - past 3 months
Question: How do you expect the demand (turnover) for your company's services to change over the next 3 months? Answer: It will (1) increase, (2) remain unchanged, (3) decrease	Evolution of demand (turnover) - next 3 months

Quarterly questions	**Topic of the question**
Question: How has your operating income changed over the last 3 months? Answer: It has (1) increased, (2) remained unchanged, (3) decreased	Evolution of operating income - past 3 months

Question: How do you expect your operating income to change over the next 3 months? Answer: It will (1) increase, (2) remain unchanged, (3) decrease	Evolution of operating income - next 3 months
Question: How have your operating expenses changed over the last 3 months? Answer: They have (1) increased, (2) remained unchanged, (3) decreased	Evolution of operating expenses - past 3 months
Question: How do you expect your operating expenses to change over the next 3 months? Answer: They will (1) increase, (2) remain unchanged, (3) decrease	Evolution of operating expenses - next 3 months
Question: How has your profitability changed over the last 3 months? Answer: It has (1) increased, (2) remained unchanged, (3) decreased	Evolution of profitability - past 3 months
Question: How do you expect your profitability to change over the next 3 months? Answer: It will (1) increase, (2) remain unchanged, (3) decrease	Evolution of profitability - next 3 months
Question: How has your capital expenditure changed over the last 3 months? Answer: It has (1) increased, (2) remained unchanged, (3) decreased	Evolution of capital expenditure - past 3 months
Question: How do you expect your capital expenditure to change over the next 3 months? Answer: It will (1) increase, (2) remain unchanged, (3) decrease	Evolution of capital expenditure - next 3 months
Question: How has your competitive position changed over the last 3 months? Answer: It has (1) improved, (2) remained unchanged, (3) deteriorated	Evolution of competitive position - past 3 months
Question: How do you expect your competitive position to change over the next 3 months? Answer: It will (1) improve, (2) remain unchanged, (3) deteriorate	Evolution of competitive position - next 3 months

B. Consumer tendency surveys

Monthly questions	**Topic of the question**
Question: How has the financial situation of your household changed over the last 12 months? Answer: It has… (1) got a lot better, (2) got a little better, (3)	Evolution of household financial situation - past 12 months

stayed the same, (4) got a little worse, (5) got a lot worse, (6) don't know	
Question: How do you expect the financial situation of your household to change over the next 12 months? Answer: 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 household financial situation - next 12 months
Question: How do you think the general economic situation in the country has changed over the past 12 months? Answer: 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 - past 12 months
Question: How do you expect the general economic situation in the country to change over the next 12 months? Answer: 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 - next 12 months
Question: How do you think that consumer prices have changed over the last 12 months? Answer: 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 - past 12 months
Question: How do you expect the consumer prices to change over the next 12 months? Answer: 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 consumer prices - next 12 months
Question: How do you expect the number of unemployed in this country to change over the next 12 months? Answer: They will...(1) increase more sharply, (2) increase slightly, (3) remain the same, (4) fall slightly, (5) fall sharply, (6) don't know	Evolution of the number of unemployed people - next 12 months
Question: 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? Answer: (1) yes it is, (2) it's neither the right nor the wrong moment, (3) no it is not, (4) don't know	Right moment to make major purchases - current
Question: 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? Answer: 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	Evolution of money spent on major purchases - next 12 months
Question: Considering the general economic situation, do you think that now it is..? Answer: (1) a very good moment to save, (2) a fairly good moment to save, (3) not a good moment to save, (4) a very	Opportunity to save - current

bad moment to save, (5) don't know	
Question: Over the next 12 months, how likely it is that you save any money? Answer: (1) very likely, (2) fairly likely, (3) not likely, (4) not at all likely, (5) don't know	Opportunity to save - next 12 months
Question: Which of these statements best describes the current financial situation of your household? Answer: 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	Statement best describing the household financial situation - current

Quarterly questions	**Topic of the question**
Question: How likely would you buy a car in the next 12 months? Answer: (1) very likely, (2) fairly likely, (3) not likely, (4) not at all likely, (5) don't know	Likelihood to buy a car - next 12 months
Question: 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.)? Answer: (1) yes definitely, (2) possibly, (3) probably not, (4) no), (5) don't know	Plans to buy or to build a home - next 12 months
Question: How likely would you spend any large sums on home improvements or renovations over the next 12 months? Answer: (1) very likely, (2) fairly likely, (3) not likely, (4) not at all likely, (5) don't know	Likelihood to spend large sums of money on home improvements or renovations - next 12 months

References

Abberger, K. and W. Nierhaus (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.

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

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., Barrell, R., and Davis, E P. (2009). *Consumer confidence indices and short-term forecasting of consumption.* Manchester School, 77(1), 96–111.

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. Available at: http://ec.europa.eu/eurostat/documents/4587857/4587935/CPP3-ALGERA.pdf

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*. Available at: http://www.aapor.org/Best_Practices1.htm.

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.

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 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

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

Basile B., S. de Nardis and A. Girardi, (2009). *Pricing to market of Italian exporting firms.* Applied Economics, vol. 41(12), 1543-1562.

Basile R., S.de Nardis and C. Pappalardo (2014). *Firm Heterogeneity and Regional Business Cycles Differentials.* Journal of Economic Geography, December 10, 2013

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.

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, 169-194.

Bovi M. (2009), Economic versus psychological forecasting. Evidence from consumer confidence surveys, Journal of Economic Psychology, vol. 30.

Bovi, M., (2013). *Are the representative agents' beliefs based on efficient econometric models?*. Journal of Economic Dynamics and Control, 37(3), 633-648.

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.

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.

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

Carlson, J.A. and Parkin, M. (1975), *"Inflation expectations",* Economica, vol. 42.

Carroll C.D. (1992). *The Buffer Stock Theory of Saving: Some Macroeconomic Evidence*. Brooking Papers on Economic Activity, 2, pp. 61-155.

Carroll, C. D. (2003). *Macroeconomic expectations of households and professional forecasters*. The Quarterly Journal of Economics, 269-298.

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.

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.

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.

Clemen, R.T. (1989). *Combining forecasts: a review and annotated bibliography*. International Journal of Forecasting, 5, 559-583

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

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.

Cicchitelli, G., Herzel, A., Montanari, G.E. (1992). *Il campionamento statistico*. Il Mulino, Bologna.

Crosilla, L., Malgarini, M. (2011), *"Behavioural models for manufacturing firms: analysing survey data"*, in Journal of Industrial and Business Economics, vol. XXXVIII – n.4, 2011 December.

Croushore, D. and T. Stark (2001). *A real-time data set for macroeconomists*. Journal of Econometrics, 105, 111-130.

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.

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.

D'Elia, E. (1991). *La quantificazione dei Risultati dei Sondaggi Congiunturali: un confronto tra procedure*. In Rassegna di lavori dell'ISCO, Istituto Nazionale per lo Studio della Congiuntura, Rome

D'Elia, E. (2005). *Using the results of qualitative surveys in quantitative analysis*. ISAE WP series n.56, September.

De Boef, S., P.M. Kellstedt (2004). *The political (and economic) origins of consumer confidence*. American Journal of Political Science, 48, 633–649.

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.

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.

De Waal, T., Pannekoek, J., Scholtus, S. (2011). *Handbook of Statistical Data Editing and Imputation*. J. Wiley & Sons, New York.

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

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.

Desroches, B. and M. A. Gosselin (2004). *Evaluating Threshold Effects in Consumer Sentiment*. Southern Economic Journal, 70 (4), 942-952.

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.

Deville, J. C., Särndal C. E. and Sautory O. (1993). *Generalized Raking procedures in Survey Sampling*. Journal of the American Statistical Association, June, Vol. 887, n:423, pp 1013-1020.

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.

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

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.

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

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

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://ec.europa.eu/eurostat/ramon/statmanuals/files/KS-BG-03-001-__-N-EN.pdf

Eurostat (2009), *ESS guidelines on seasonal adjustment*. Available at: http://ec.europa.eu/eurostat/documents/3859598/5910549/KS-RA-09-006-EN.PDF/0a9893a6-3e69-45c4-8328-6fc5a3497c13

EUROSTAT, *Short term business statistics* Available on: http://epp.eurostat.ec.europa.eu/cache/ITY_SDDS/en/sts_esms.htm.

EUROSTAT (2010). *The European Business Cycle Clock*. Available at: 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.

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.

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.

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.

Garner, C. A. (2002). *Consumer confidence after September 11*. Federal Reserve Bank of Kansas City, Economic Review 2, 1–21.

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).

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.

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

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.

Hartley, H. O. (1974). *Multiple-frame methodology and selected applications*. Sankhya, Series C, 36, 99–118.

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.

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.

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

ILO (2013). *Measuring informality: A statistical manual on the informal sector and informal employment*. Available at: http://www.ilo.org/wcmsp5/groups/public/---dgreports/---dcomm/---publ/documents/publication/wcms_222979.pdf

IMF, SDDS Guide Available on: http://dsbb.imf.org/pages/sdds/home.aspx.

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

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.

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

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.

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.

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, State University of New York, Albany, New York, 12222, USA, April.

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.

Lohr, S., & Rao, J. N. K. (2006). Estimation in multiple-frame surveys. Journal of the American Statistical Association, 101, 1019–1030.

Lothian, J. and Morry. M. (1978), *A set of Quality Control Statistics for the X-11-ARIMA Seasonal Adjustment Method*, Working Paper 78-10.

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.

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

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.

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.

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.

Matsusaka, J. G. and A. M. Sbordone (1995), *Consumer Confidence and Economic Fluctuations*, Economic Inquiry, vol. XXXIII, n.2, pp. 296-318.

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.

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 (2008). *Handbook on Constructing Composite Indicators. Methodology and user guide*. OECD Publications, Paris, ISBN 978-92-64-04345-9. (http://www.oecd.org/els/soc/handbookonconstructingcompositeindicatorsmethodologyanduserguide.htm)

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.

OECD. *Leading Indicators and Tendency Surveys Portal*. http://www.oecd.org/std/clits/

OECD. Main Economic Indicators. http://stats.oecd.org/Index.aspx?DataSetCode=MEI_BTS_COS

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

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.

Ramalho, E. A., A. Caleiro and A. Dionfsio (2011). *Explaining consumer confidence in Portugal*. Journal of Economic Psychology, 32, 25-32.

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.

Roos, M. W. (2008). *Willingness to consume and ability to consume*. Journal of Economic Behavior & Organization, 66, 387–402.

Rottman H. and T. Wollmershauser (2013). *A micro data approach to the identification of credit crunches*. Applied Economics, 47 (17), 2423-2441.

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.

Särndal, C.E., Lundström, S., (2005). *Estimation in Surveys with Nonresponse*. J.Wiley & Sons, New York.

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. Available at: http://www.insee.fr/fr/methodes/outils/calmar/doccalmar.pdf

Sautory, O. (1993). *CALMAR (Calage sur marges).* Available at: http://www.insee.fr/fr/methodes/default.asp?page=outils/calmar/accueil_calmar.htm

Skinner, C. J., Holmes, D. J., & Holt, D. (1994). Multipleframe sampling for multivariate stratification. International Statistical Review, 62, 333–347.

Seitz, H. (1988), *The Estimation of Inflation Forecasts from Business Survey Data*, Applied Economics, 20(4), 427-438.

Shiskin, J., A. H. Young, and J. C. Musgrave (1967). *The X-11 variant of the Census Method II seasonal adjustment program,* Technical Paper No. 15, U.S. Department of Commerce, U. S. Census Bureau.

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.

Statistics Netherlands, 2011, Method series 11, *Seasonal adjustment*,(http://www.cbs.nl/NR/rdonlyres/75B108E9-FE3E-4506-904D-166AD9F785A6/0/2011x37062art.pdf).

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 a large number of predictors.* Journal of the American Statistical Association, 97, 1167-1179.

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.

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.

Throop, A. W. (1992). *Consumer sentiment: its causes and effects*. Economic Review, 35–59.

United Nations (1986). *National Household Survey Capability Programme: Sampling Frames and Sample Designs for Integrated Household Survey Programmes (Preliminary Version*). New York. Available at: http://unstats.un.org/unsd/publication/unint/DP_UN_INT_84_014_5E.pdf

UNECE (2015) *Guidelines on Statistical Business registers* Available online at: http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/2015/3_Add.1_Rev1_Guidelines_on_BR_3.06.2015.pdf

UNSD (2005) Household Sample Surveys in Developing and Transition Countries. Studies in Methods, Series F No. 96. Available at: http://unstats.un.org/unsd/hhsurveys/

UN Statistics Division (2009), "*International Recommendations for Industrial Statistics 2008"*, New York, 2009, available at http://unstats.un.org/unsd/industry/docs/M90.pdf

University of Michigan (2013). *Survey of consumers*. Accessed on May, 2013 http://www.sca.isr.umich.edu/documents.php?c=i

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).

U.S. Census Bureau (2013), *X-13-ARIMA SEATS Reference Manual*, *Version 1.1,* Time Series Research Staff, Statistical Research Division, U.S. Bureau of the Census, Washington D.C.

Vanderhoeft, C. (2001). *Generalised Calibration at Statistics Belgium*. Statistic Belgium Working Paper N. 3, NIS. Available at: http://www.statbel.fgov.be/studies/paper03.pdf

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.