Industrial Statistics

Guidelines and Methodology

UNIDO

UNITED NATIONS
INDUSTRIAL DEVELOPMENT ORGANIZATION

United Nations Industrial Development Organization

Industrial Statistics

Guidelines and Methodology

Vienna
2010

The description and classification of countries and territories in this publication and the arrangement of the material do not imply the expression of any opinion whatsoever on the part of the Secretariat of the United Nations Industrial Development Organization concerning the legal status of any country, territory, city or area, or of its authorities, or concerning the delimitation of its frontiers or boundaries, or regarding its economic system or degree of development. The designation "country or area" covers countries, territories, cities and areas. The designation "region, country or area" extends the preceding one by including groupings of entities designated as "country or area". Designations such as "industrialized" and "developing" countries and "newly industrialized countries" are intended for statistical convenience and do not necessarily express a judgement about the stage reached by a particular country or area in the development process.

Table of Contents

A Business Register is an essential tool for industrial statistics. It serves as a statistical frame for Industrial Surveys and as an important tool for managing an industrial census and controlling for under coverage. A system for updating the Business Register is critical for the preparation of reliable time series for employment, value added and business demography in the industrial sector. This chapter provides a detailed description of the data sources of a Business Register (including how to identify new entries and the use of a matching technique to avoid duplication of entries), the data structure of a Business Register and the updating process and updating system.

This chapter provides a detailed description of the data items for inclusion in Industrial Surveys. The chapter begins with a brief description of the survey methodology proposed, followed by a presentation of the questionnaire design and coverage of data items for the different purposes of Industrial Surveys. Definitions of all the data items used in these questionnaires are also given. Both, large and small questionnaires are presented.

This chapter describes different methodologies that can be considered for developing the sampling frames and designing the samples for Economic Surveys. Procedures for updating the sampling frames are also described. These guidelines are mostly designed for statisticians working with establishment surveys in developing countries, but others may benefit from the discussion of the various conceptual issues. These guidelines are designed to be practical, with examples of applications in different countries, and has limited discussion of statistical theory; hence only an overview of the basic sampling concepts is presented.

Acknowledgements

Preface

Technical cooperation (TC) with the National Statistics Offices (NSOs) of developing countries and countries with economies in transition has been a matter of prime importance for UNIDO since the very beginning of its statistical activities in 1979. The first UNIDO technical cooperation programme began in the second half of the 1980s in response to the 1983 World Programme of Industrial Statistics. This was the third in a series sponsored every 10 years by the United Nations Statistical Commission, the main aim of which was to encourage the orderly development of national enquiries into the structure and activity of the industrial sector

The overall aim of the UNIDO TC programme was to contribute to the World Industrial Statistics Programme by improving the capacity of developing countries to collect, process, use and disseminate industrial statistics. In the framework of UNIDO's overall technical assistance programme, industrial statistics has been regarded as part of the 'upstream' activities, supplying industrialists and policy makers with the information needed to make coherent assessments and effective decisions. As technical assistance in industrial statistics expanded and the number of projects grew, UNIDO developed a broader, more encompassing approach which established the conceptual framework of the National Industrial Statistics Programme (NISP) in 1989.

NISP attracted widespread interest among developing countries. At the start of the 1990s, as a strong wave of economic liberalisation and privatisation hit many countries with centrally planned economies, industrial statistics systems faced new challenges. Earlier conventional forms of data collection had assumed that data users would essentially be government officials. However, in the new emerging situation, industrial data were also required by private sector actors, production associations and potential domestic and foreign investors. Therefore, NISP placed high priorities on users' need and their ability to access data quickly. NISP projects involved national chambers of commerce and industry and other business membership organisations in discussions on the survey questionnaire and in the supervision of data collection. The publication of the business directories containing the most recently updated information for the wide use of the private sector became one of the major outputs of the project. After some two decades, NISP was discontinued as the new priorities of industrial statistics emerged in many countries.

The industrial statistics programme in the new context is not confined merely to basic data collection from a field survey. It supports the compilation of a well-maintained and regularly updated Business Register with the administrative records and survey data, a comprehensive data set produced from periodic censuses and regular annual surveys, and a

system of short-term indicators including the index of industrial production compiled from monthly or quarterly surveys. In response to the demand for an entirely new system of industrial statistics, in its recent sessions, the United Nations Statistics Commission has approved a series of new recommendations and standards, which include:

- International Standard Industrial Classification (ISIC) of all economic activities, revision 4, 2006 (previously, ISIC revision 3, 1990)

- International Recommendations for Industrial Statistics (IRIS), 2008 (previous recommendation was adopted in 1983)

- The System of National Accounts (SNA), 2008 (previously, SNA 1993)

- International recommendations on index numbers of industrial production, 2010 (previous manual dated back to 1950)

These recommendations basically determine the methodological uniformity and international comparability of industrial statistics. The new recommendations also adequately address the recent changes of priority in industrial statistics. Currently, UNIDO's technical cooperation programme is directed at the implementation of the new standards in the national statistical system. In a number of countries, UNIDO is introducing new standards through ongoing technical assistance projects. In many others, the UNIDO Statistics Unit has been actively involved in several international workshops and training programmes aimed at building awareness in NSO statisticians of the fundamental changes made in the new standards. Many NSO statisticians have also emphasised the need for a single reference manual that synthesises the entire conceptual framework for regular Industrial Surveys.

This publication is intended to serve as a handbook for statisticians involved in the regular industrial statistics programmes of NSOs or line ministries. It describes the statistical methods related to the major stages of industrial statistics operation. These include:

- maintenance and updating of the Business Register, which provides the frame for an Industrial Survey *(Chapter 1)*

- data items for the survey of large and small establishments *(Chapter 2)*

- methods of data processing *(Chapter 4)* and

- data analysis based on a set of statistical indicators of industrial performance *(Chapter 6)*

These guidelines also include a chapter on practical guidelines for sample design for Industrial Surveys *(Chapter 3)*. A separate chapter is dedicated to the short-term

indicators of industrial statistics and this also covers the construction of indices of industrial production from the monthly survey data. It also includes a description of the Prima software for integrating production and producer price indices *(Chapter 5)*.

Concepts and methods described in the publication are based on the recent UN recommendations and standards, as well as the best national practices related to industrial statistics. The model questionnaires presented in the publication meet the data requirement not only for compilation of principle indicators of official industrial statistics in the overall framework of macroeconomic accounting but also key variables of business trend statistics. However, this does not imply that each Industrial Survey should include all data items presented in this publication. Statisticians of NSOs make their own selection of data items and survey methods depending upon the data requirement of the country at a particular time. The authors of this publication are aiming to provide the reference materials for the key aspects of a national industrial statistical system.

Shyam Upadhyaya

Chief Statistician

United Nations Industrial Development Organization

Acronyms

API	Application Programme Interface
ASI	Annual Survey of Industry
CIP	Competitive Industrial Performance
CLI	Common Language Infrastructure
COSIT	Centre of Statistics and Information Technology
CPC	Central Product Classification
CPI	Consumer Price Index
CV	Coefficient of Variation
DEFF	Design Effect
EA	Enumeration Area
FIRST	Fully Integrated Rational Survey Technique
GDP	Gross Domestic Product
HDI	Human Development Index
HIES	Household Income & Expenditure Survey
ICT	Information and Communication Technology
IDSB	Industrial Demand Supply Balance database
IIP	Index of Industrial Production
IMF	International Monetary Fund
IMPS	Integrated Microcomputer Processing System
IRIS	International Recommendations for Industrial Statistics
ISIC	International Standard Industrial Classification
MHT	Medium-High Technology
MIS	Management Information System
MLI	Matching Likelihood Index
MSE	Mean Square Error
MVA	Manufacturing Value Added
NISP	National Industrial Statistics Programme
NSO	National Statistics Office
ODBC	Open Data Base Connectivity
PDA	Portable Digital Assistant
PPI	Producer Price Index

PPS	Probability Proportional to Size
PRIMA	PRice (or PRoduction) Index MAnagement System
PSU	Primary Sampling Unit
SITC	Standard International Trade Classification
SNA	System of National Accounts
SO	Statistics Office
SQL	Structured Query Language
SSU	Secondary Sampling Units
UN	United Nations
UNDP	United National Development Programme
UNIDO	United National Industrial Development Organization
UNSD	United Nations Statistical Division
VAT	Value Added Tax
WPI	Wholesale Price Index
XML	Extensible Markup Language

Chapter 1: Business Registers for Industrial Statistics

1.0   Introduction

A Business Register is an essential tool for industrial statistics for three major reasons. First, it serves as a statistical frame for Industrial Surveys and as a basis for grossing up results from sample surveys in order to produce business population estimates. Second, it serves as an important tool for managing an industrial census and controlling for under coverage. Third, a system for updating the Business Register is essential for the preparation of reliable time series for employment, value added and business demography in the industrial sector, as will now be explained.

A Business Register is a list (usually in the form of a computer file) of the enterprises and/or establishments in a country, with an identification number for each unit. It includes information about the unit's size, kind of activity, and activity status, as well as extensive contact information.

Each year, some industrial establishments cease operations while others commence operations; a register that does not adequately reflect these changes will soon become obsolete. In developing countries with substantial industrial growth, births tend to exceed deaths, so that the true number of establishments tends to grow from year to year. A Business Register that is not properly updated may fail to keep up with growth in the number of establishments and enterprises. Many countries lack systems for updating their Business Registers consistently between censuses; accordingly, time series for employment and value added do not reliably track annual industrial growth.

The *Business Register: Recommendations Manual* of the European Communities stresses the importance of good business demography, that is, statistics of the births and deaths of establishments.[1] For many countries, however, industrial demography is unreliable. A census of industry, if based on door-to-door canvassing, may lead to the discovery of many establishments that were missed during updating. In that case, time series for the numbers of establishments and persons engaged might show a sharp increase during census years and little or no increase during inter- and post census years. If that were to happen, the time series for these aggregates would not reflect changes in industrial activity but would instead reflect changes in the activity of the statistics office. The example cited above illus-

---

[1] European Communities, "Business Register: Recommendations Manual, 2010 Edition", Office for Official Publications of the European Communities, Luxembourg. See:
http://unstats.un.org/unsd/EconStatKB/KnowledgebaseArticle10171.aspx?Keywords=Eurostat

trates the need for a consistent system for updating the register from year to year so that changes in it will reliably reflect changes in the industrial sector.

As regards the register, the UN Recommendations for Industrial Statistics (IRIS) 2008 differ from the previous version published in 1983 in these major respects:

- They recommend reliance on a comprehensive Business Register, whereas the previous version left open the option of a separate industrial directory. The comprehensive register is preferred because it ensures consistency in the treatment of the various activities and because it is more efficient for a single unit within the National Statistics Office to prepare the Business Register than for various units to prepare separate registers.

- It is recommended that the statistical units for the R register be both establishments and enterprises, with links to each other, whereas the 1983 version recommended that the units be establishments including central offices, with links to central offices in the case of multi-establishment enterprises. This in turn reflects the difference in data collection approaches: the 1983 version treats the establishment as the basic statistical unit for data collection whereas the new recommendation is that the establishment should be treated as the basic unit for collection of production data only, with the enterprise serving as the basic unit for collection of data by institutional sectors, as required for the System of National Accounts 1993 (SNA93).

- IRIS 2008 focuses for the first time on the strategy for data collection. In that context, it stresses the importance of using administrative data and mentions that: "Countries with developed statistical systems make progressively more use of administrative sources for coverage of industrial activities". The 1983 Recommendations focused solely on the collection of data by means of surveys and censuses.

- IRIS 2008 refers to the practical need in many countries for a cut-off that limits register coverage to units of more than a certain size (usually employment size) in order to limit the size of the register and the cost of updating. This need is particularly felt in developing countries due to the relatively high cost of register updating in those countries and the large numbers of small industrial establishments.

2.0   Ancillary Uses of the Business Register

Discussion of the Business Register would not be complete without mention of two ancillary uses for the industrial data contained in a well ordered register.

Using the Business Register as a database, a customised application can be prepared that can greatly assist in the implementation of an Annual Survey of Industry (ASI) or any other

industrial survey. For example, in Indonesia, a module for the register updating system prepared in the early 1990s tracked receipts of ASI questionnaires, as well as the transmission of the questionnaires to the central office in Jakarta. The programme was designed to print out automatically a manifest of all the questionnaires being sent to Jakarta – a task that had previously been executed by typewriter. It could also print a list of establishments for which questionnaires had not yet been received. The convenience of these features helped motivate the provincial statistics offices to put more effort into cleaning up their own industrial registers. At a later stage, a separate application was developed for use at the district level. In addition to tracking receipts and transmittals of ASI questionnaires, it also tracked the response rate for establishments managed by each enumerator and identified uncooperative establishments which were then visited by supervisors. With the help of this application, the statistics office was able to push non-response down to very low levels.

Similarly, in Sri Lanka during 2006/2007, a module was developed for the register updating system that tracked the progress of efforts to obtain a questionnaire from each ASI establishment. This included support for a call centre that placed calls to establishments that had not yet responded to the ASI. The module is still a prototype and has not yet been applied operationally.

Another ancillary use of the industrial register is publication, where this does not contravene the confidentiality provisions of the statistics law. An industrial register of about 20,000 establishments has been published in Indonesia and this has also motivated provincial offices to clean up their own registers in preparation for publication. In Sri Lanka, however, publication has not yet been possible because of concern that it would contravene the confidentiality provisions of the statistics law.

3.0 Data Sources

Business Registers are created and updated using various data sources. The availability of the sources and their characteristics differs from country to country but some general features can be noted here. Moreover, although not discussed in detail in this chapter, other data sources, such as trade/shipping manifests, newspapers, local authority registers, trade directories and Yellow Pages, electricity and other utilities, population and housing censuses may also be used.

3.1 Data Sources for Entries

3.1.1 Tax registers

Where available, tax registers provide the optimal data sources for the Business Registers. Lists for Value-Added Tax (VAT) are often used but other tax lists would also be suitable. It

should, however, be noted that tax registers vary in regard to taxable units. Registers for payroll taxes (such as those for social security or unemployment insurance taxes) treat establishments as the taxable unit, while others treat enterprises as the taxable units. A tax register will normally include basic identifying information such as name, address, geo-code, telephone, e-mail address and fax numbers, and some indicator of size.

Tax registers are suitable for two reasons. First, they are fairly comprehensive, thanks to ongoing tax enforcement efforts. Second, tax registers provide indicators of size and activity status, information that is often lacking in other sources. In particular, the registers can identify units that are closed or inactive, thereby eliminating the need for two burdensome tasks in register updating using other data sources. Namely, there is:

- No need to field check whether a "candidate" for addition to the Register is in operation or not

- No need to use field checks to identify which establishments have closed

For example, improved access to the VAT register led to a sharp improvement in the completeness of the South African Business Register in 1991.[2] Tax data are commonly used as a basis for register updating in Europe, North America, Australia and New Zealand. They are also often used for updating in transition on countries but unfortunately such data are often unavailable to statistics offices in many developing countries.

3.1.2 A single, non-tax administrative source

Many countries do have a register of companies but this would only be suitable for updating a statistical register if it contained information on economic activity, employment, a current establishment address and activity status. Some countries require new enterprises to register with the National Statistics Office, providing an opportunity for that office to obtain some of the information required for register updating. In such a system, however, the statistics office typically lacks a method for documenting closures with the result that, over time, closed enterprises account for a larger and larger proportion of the register.

In many countries, the register of companies is complete but lacks essential information and cannot serve for updating a statistical register until it is restructured in support of that purpose. For example, it would be essential to collect information on the address of the production facility, the main product, and the number of employees. In this case too, as in the case of the tax register, enhanced cooperation between government agencies is needed to improve the Business Register.

---

[2] Arrow, Jairo and Nilsson, Gösta, "Progress Report" for the 14th International Roundtable on Business Survey Frames, Auckland, 2000.

3.1.3 Multiple administrative sources

In the absence of a single source, a Business Register can still be updated using multiple sources, with each source providing information about some of the establishments that are missing from the register. An effective updating system based on multiple sources will require teamwork by a small group of skilled staff who understand the system and can make purposeful decisions. When multiple sources are used, however, the updating task is more complicated and will provide a less satisfactory solution, for several reasons:

- Because the multiple sources will, typically, lack a common identifier, the statistics office will have to match the various sources against each other and the statistical register in a laborious procedure that is subject to error[3]

- Duplicate entries may arise for one and the same establishment because of matching errors

- The statistical register cannot be expected to provide complete coverage since none of the sources provides complete coverage

- In the absence of tax data, no source may be available to document closures

- Field checks, which are costly, will be needed to confirm whether the establishment qualifies for the register

3.1.4 Economic Census

 It is commonly believed that an economic census, involving door-to-door canvassing, provides a complete list of establishments, albeit at high cost. However, the matter is not as simple as that. The 2008 Recommendations do introduce a note of caution, mentioning that this approach will miss "the non-recognisable places of business and enterprises without fixed location". In addition, experience in several developing countries has shown that door-to-door canvassing can miss a substantial share of recognisable establishments for reasons that will be discussed in section 8. In view of its high cost, a census is obviously not a practical source for identifying new establishments on an annual basis.

3.2.   Data Sources for Exits

The documentation of exits is important in a different way from that of entries. For a business surveys frame, a failure to document an entry means that a unit will be missed from the universe, while a failure to document an exit does not lead directly to coverage error in the survey. If proper survey procedures are followed, the finding that a unit has

---

[3] Steven Vale, "The Use of Administrative Sources for Economic Statistics – An Overview", paper delivered at Regional Workshop for CIS on the Use of Administrative Data for Economic Statistics, Moscow, 2006.

closed can easily be accommodated in the blow up of results to the universe. Nevertheless, a register that is cluttered with a large number of inactive units is unsatisfactory for three reasons:

- It will confuse enumerators when they go to the field and may lead to an increase in the number of cases that "cannot be found"

- It will impair the efficiency of samples based on criteria such as geo-code and size

- It cannot serve for business demography unless adjusted for exits based on the findings of sample surveys

The strengths and weaknesses of the various data sources are quite different for exits than for entries. A very effective source for exits, but of no use for entries, is feedback from Industrial Surveys, especially for an ASI with complete coverage (no sampling) above a certain threshold.

Unfortunately, many countries still fail to utilise this source for documenting exits even though it is available at virtually no cost for countries that rely on enumerator visits to conduct an ASI. In many countries, the statistics office merely notes which units have or have not responded to the ASI but does not enquire systematically into the reason for failure to respond – for example, whether the establishment has closed, or fallen out of scope, or is still active but simply failed to fill out the survey form. What is needed is simply a null response form for the enumerator to fill out in case of inability to obtain a completed survey questionnaire from the establishment. In principle, two kinds of null returns may be used:

- When ASI questionnaires are delivered by enumerators, as is the case in many countries, a null return should be filed at once by enumerators for all units to which questionnaires cannot be delivered. This may be because the unit is permanently closed, temporarily closed, out of scope, or discovered to be duplicate or merged with another unit.

- At units where ASI questionnaires have been delivered by enumerators but for which a response is not forthcoming by the deadline, a null return should be filed by enumerators. This form will provide positive confirmation that the unit is still active but has not responded, or will mention another reason for the non-response if the unit is no longer active or has fallen out of scope. For active non-respondents, it should also be possible to obtain an employment estimate from security guards or neighbours.

In practice, some countries have preferred to use a single null return to cover both cases. The form covers all of the above-mentioned reasons for lack of a completed ASI

questionnaire. A single null return is particularly suitable in the case of countries for which questionnaires are initially sent out by mail so that an enumerator visit, if there is one, takes place only after the unit has failed to respond to the request by mail.

Even if part of the ASI universe is covered on a sample basis, the ASI can still provide good coverage of exits if the sample is changed and is rotated from year to year to assure coverage of all registered establishments once in two or three years, or even once in five years as is done in India.

A tax register is an excellent data source for exits, as it will show when units cease to have income or payroll or fall out of the register. Another fairly good source is a list of industrial or other customers of electric utilities but this is subject to some error inasmuch as consumers may continue to consume electricity under an industrial tariff even after an establishment has closed. Many other administrative data sources are unsuitable for recording exits, since they do not monitor the ongoing activity of registered units.

In some countries, the Business Register is updated by means of a special survey specifically for that purpose. In some advanced countries, this survey is carried out by post – an approach that is feasible wherever businesses have a good record of responding to postal inquiries. In many transition countries, a Business Register survey is carried out by field visits or post; response rates are very high due to the tradition of compulsory reporting by establishments. In developing countries, however, postal surveys often have very low response rates.

A well-ordered Business Register includes records for enterprises and establishments that were previously active as well as for ones that are currently active. Accordingly, records for exiting units should, under no circumstances, be deleted from the Business Register. Exits may occur for many reasons and a Business Register needs activity status codes for the various reasons: closed permanently, closed temporarily, out of scope, duplicate, etc. When duplicates are discovered, it is important to inactivate them in a systematic way. In some countries, a problem has been reported for establishments that cannot be found during ASI field enumeration. The statistics office may be reluctant to treat the establishment as closed until closure can be confirmed. A "not found" category is an option but the statistics office may fear that this category could be abused by enumerators who fail to look for or to enter the unit. In the absence of such a category, however, the enumerators will be puzzled as to how to classify an establishment that they cannot locate in the field.

3.3 Census of Establishments

With regard to the role of a census of establishments or economic census in capturing entries and exits, it is commonly believed that a census is a statistical "gold standard" for populating a Business Register and that its only drawback is its great cost particularly in the

case of annual updating. In fact, however, the coverage of a census cannot be assumed to be complete.

IRIS 2008 mention that a census will miss "the non-recognisable places of business and enterprises without fixed location". This is true, yet another problem is that many recognisable places are missed by enumerators for more mundane reasons – such as the following:

- Difficulty to enter establishments where security guards are told to keep out visitors. In this case, the statistical agency may have the legal authority to enter and deliver questionnaires but may, in practice, lack the institutional authority to utilise its legal authority.

- Census enumerators, who are typically temporary workers with little or no previous experience with the statistical agency, may occasionally ignore recognisable places for reasons of their own, in violation of their instructions, while supervisors, who may be permanent employees of the statistical agency, may lack the tools to monitor compliance with the instructions. The risk of this kind of failure is particularly great in large cities, especially the capital city, where the number of establishments is very great and the resources devoted to the canvassing are not always adequate.

The omission of recognisable establishments by census enumerators has been observed in several developing countries when the census results are compared with a Business Register or a reliable administrative list. The basic problem seems to be the lack of a good control mechanism for use by supervisors. A well-maintained Business Register could provide a basis for control but is often neglected as a starting point for an establishment census.

Countries that do update their register between censuses face special technical issues in managing the integration of the two sources when they conduct a census:

- The preferred, but less frequently used, technique is to carry the list of establishments in the register to the field during enumeration. This will ensure a good linkage between the census listing and the old register so that the census leads to new discoveries without losing any of the information in the old register. In addition, this approach will identify exits systematically. This method is costly, however, in management terms. It requires that the old register be accurately sorted by small areas and that enumerators should be well trained as to how to handle possible discrepancies in the field.

- The common, but less appropriate, technique is to go to the field with a "blank slate," that is, without the old register. In this case, supervisors will lack a control

mechanism for checking on the completeness of enumeration. Accordingly, the census will produce a new list that should, in principle, be more complete than the old register but which, in practice, often misses out many establishments in the old register. After the census, the statistics office will face a dilemma: Ignore the old register or combine it with the census? If ignored, much useful data, including longitudinal data for specific establishments (via the link between an old Establishment Identification Number (EIN) and a new one), will be lost. If combined, headquarters will face a messy post census job of matching the new list against the old register and one that will also be costly in management terms.

4.0 Alternative Sources for Identifying New Establishments

The first two sources mentioned in Section 3.1 – a tax register and a single, non-tax administrative list – greatly facilitate the task of identifying new establishments and enterprises. If the Business Register relies entirely on one or the other of these sources, there will be a common identifier as between the Business Register and the external source and it will be easy to identify units that have been added to the external source as well as to ascertain whether the unit is or is not already in the Business Register.

The next question is whether the unit, if not currently in the Business Register, should be added to it and, if so, what further information is required. For entries from a tax register, recent activity status will be clear from the tax data but it may still be necessary to contact the establishment to obtain more complete information about employment size, kind of activity, and contact details such as telephone numbers and contact person. For entries from a single, non-tax administrative list, it may be necessary to inquire about all of these matters plus the activity status (some enterprises may have registered but never become active).

The case of multiple administrative lists is more complicated. For new establishments, the updating system requires the annual availability of external lists of establishments or enterprises from administrative sources, based on self-registration with those sources. For effective targeting of new establishments, it is preferable that the external lists include only establishments that are active, provide indicators of size and economic activity, and up-to-date names, addresses, telephone numbers, and e-mail addresses. It is rare for external sources to provide all the required data but available lists with partial data may be useful even if they do not fully conform to these criteria. Examples of the suitability of various lists are mentioned in Annex 1 describing the Sri Lankan experience.

Some countries combine data from multiple sources with data from the field – either from casual observation by enumerators or from a census of establishments.For the multiple

source case, the system for identifying new establishments involves six procedures in sequence:[4]

1. *Importing, parsing and editing* of the external lists into an integrated system. Parsing serves to put names, addresses and phone numbers, etc., into a common format. Further editing (with follow-up phone calls where necessary) may be required to assign geo-codes and clarify questions about the address.

2. *Matching* serves to identify establishments in the external sources that are already found in the core register or in another external source.

3. A list of *unduplicated candidates* for addition to the register is prepared by consolidating all establishments from the external lists that are unmatched to the core register. The candidates are said to be unduplicated because, in the case of matched listings in the various external lists, only one is selected.

4. In preparation for field checks, the *candidates are prioritised* based on various indicators (source, size, year in which production started, etc.) that are believed to be correlated with the likelihood of qualifying for the register. High priority candidates are designated for field checks; medium priority candidates may be designated for phone checks; while low priority candidates may be checked on a sample basis or not be checked at all because of lack of resources.

5. *Field and phone checks* are carried out and the data are entered for both successful and unsuccessful candidates.

6. The final step, *copying successful candidates to the register,* can take place once the successful candidates have been vetted for possible duplication with the core register.

This system, based on multiple sources, has been tested and applied in both Sri Lanka and Indonesia and found to be effective in discovering large numbers of establishments that are missing from the core register. But the use of multiple sources complicates the updating task. Much judgement is required at each step of the way: selection of external data sources; resolving ambiguous match cases; and selection of priority candidates for field checks. The system requires constant monitoring by management and is therefore costly in management terms, considerably more so than would be a system based on a single source. There is also the risk that, should management attention falter, the system may cease to operate properly.

---

[4] Alex Korns, "Updating the Sri Lankan Register of Industry and the Role of IT", paper for the third International Conference on Establishment Statistics (ICES), Montreal, 2007. Alex Korns, "A Workable System for Updating Indonesia's Manufacturing Directory", paper for the first ICES, Buffalo, 1993.

5.0 Matching Techniques

A system based on multiple administrative sources will require extensive matching to prevent two kinds of duplication: the field checking of units that are already in the register or in another source; and the introduction of duplicate units into the register. Matching is greatly facilitated by software that can identify likely matches. In this context, it is useful to distinguish "matching" and "blocking". A match exists when records from two different lists refer to the same establishment or enterprise; blocking is "a procedure for linking records in one list with likely matches in the same list or in one or more other lists."[5] For very large registers, commercially available matching programmes can be purchased but, for smaller registers, another option is to include a set of basic matching routines in a customised application for register updating, as discussed in section 14.

False matches are more to be feared than missed matches and the matching procedure should reflect this principle. A false match will lead to a unit not being added to the register (or not being field checked for possible addition to the register), whereas a missed match simply leads to wasted effort in field checking a unit that is already in the register, or, at worst, to a duplicate entry in the register. Duplicate entries, when discovered, can be purged using a procedure discussed in the next section.

Prior to matching, it is essential to parse and edit the data from external sources in order to convert it into a format that is suited for matching. In Sri Lanka, this involved parsing each address into eight components – street name, street number, building name, floor number, village or ward, city, postal code and a comment field. Names were parsed so as to separate the suffix (such as Ltd.) and phone numbers were parsed into an area code and local number.

The following methods have been found useful in matching:

- The heart of the system for record matching for character strings is an algorithm based on analysis of bigrams – pairs of sequential letters in a character string. Similarity of an establishment name or street name – despite small spelling variations – leads to a high bigram score and thence to a high score on a matching likelihood index (MLI).

- It is useful to assign a high MLI whenever there is agreement on the name, the address, or any single telephone number, although the other two variables may show little or no similarity. This gives the benefit of the doubt to possible matches -- for review by operators. For the address, confusion may arise due to different ways of writing the address for the same location or to inconsistencies in geo-coding or in

---

[5] William Weeks, "Register Building and Updating in Developing Countries". Paper for the third ICES, Montreal, 2007.

recording addresses for a factory and a central office. Considering the frequency of such data entry errors, it is more appropriate to allow the operator to try to sort this out than to let the programme make the decision.

- It is useful to give operators a means of marking non-matches as well as matches so that pairs of blocked units that do not match can be excluded from future consideration. However, the system must also give operators a way of reviewing such cases, as needed.

- Operators review each case with a high MLI and decide whether it is a match, a non-match, or a pending case for further analysis. Pending cases are printed out for follow-up by telephone. As proposed by Weeks, the programme for Sri Lanka was structured in such a way as to facilitate processing of cases with the highest MLIs first in order to remove these cases from further consideration. The programme was also designed to enable supervisors to review staff decisions before they were implemented.

- Final matching decisions should be made by staff, as the evidence is not always clear enough to permit decision making by a computer programme. In uncomplicated cases, staff can quickly ratify the match. In complicated cases, phone calls can help clarify ambiguous evidence.

6.0 Processing of Candidates for Addition to the Register

Using a tax register, the statistics office can simply note which units have been added since the last Business Register update and add the new unit from the tax register to the Business Register, with the assurance that the unit is not duplicated in the Business Register and is probably active. It is possible that the tax register may simply become the Business Register but there may also be some further screening to decide whether the unit belongs in the register (e.g., whether it is above a size cut-off) as well as some discovery of additional units from other sources that may not be covered by a tax register, such as non-government organizations. It may also be necessary to check employment size and the main product with a phone call or field visit.

Similarly, if the Business Register relies on a single non-tax administrative list, the statistics office can identify units that have been added to the source list since the last Business Register update and can treat these units as candidates for addition to the register. Again, some further screening may be needed prior to addition to the register. It may be necessary to check the activity status and employment size, as well as the main product, with a phone call or field visit.

After the matching of lists from multiple sources, there still remains a set of records from external sources that do not match with the Business Register. By eliminating duplication between the external sources, this can easily be converted into the unduplicated set of records that do not match the Business Register. These may be referred to as the set of "candidates for addition to the Business Register". With most such non-tax administrative sources, it is necessary to field check the "candidates" as data from the external source are insufficient to confirm that the candidate is active and belongs in the Business Register or to provide all the information required for entry into the Business Register.

As the total number of candidates is quite large when multiple sources are used, it is often too costly to field check all of them. In most cases, it is therefore useful to stratify the candidates, based on criteria that are believed to be well correlated with the probability of qualifying for the register. Typical stratification criteria would include the source (some sources have much higher success rates than others), the year of registration in the external source and an indicator of size. The highest priority stratum would be field checked, the next stratum would be phone checked for all cases for which telephone numbers are available, and the lowest priority stratum would either not be checked or would be checked on a sample basis, in which case, qualifying candidates would be copied not into the main Business Register but into a supplemental frame. The supplemental frame could be used for representing at least a part of the universe of units missed by the Business Register.

Whenever field checks are required, the results should be entered into the updating system. The system needs to provide a place for entering field data for all candidates – whether or not they qualify for the Business Register. The data for non-qualifying candidates will be useful in following years to complete documentation of the external data sources and avoid rechecking the same candidates in future years. Records for qualifying units will, of course, be copied to the Business Register.

7.0   Summary of Options for Creating and Updating a Business Register

There are various options facing countries, particularly developing countries, for the creation and updating of a Business Register and they are outlined in Table 1.

The optimum source for creation of a Business Register and for monitoring both entries and exits is a tax register (Case 1 in the table). However, making this register available for statistical purposes requires a high level of inter-agency cooperation within a national government, as well as a high level of discipline within the statistics office (to avoid breaches of confidentiality). Unfortunately, access to the tax register is rarely available to statistics offices in developing countries but it is widely used in advanced and transition countries.

Next in order of desirability is a single non-tax administrative list. Such a list can be very useful if it offers good coverage and includes essential data items such as an exact address for the production facility, an indicator of size, and an indicator of the kind of activity at the unit. Where inter-agency cooperation is strong, the agency with the administrative list may be willing to add a limited number of questions for statistical purposes. For example, the statistics office in India uses the lists of factories registered with the Department of Health and Safety (DHS, formerly the Chief Inspectorate of Factories) in each state. The use of a single source precludes any need for matching. However, even a very complete administrative list may fail to document exits so another source may be needed for that purpose. Such a mixed system is shown as Case 5 in Table 1. Another possible source is the register of companies which, in theory, has the potential to serve as a basis for the Business Register but is rarely used due to the lack of statistically-relevant data in the databases compiled by the Register of Companies.

Table 1.   Conspectus of Updating Systems for the Business Register

	Source for creation	Source for entries	Source for exits	Comment
	(A)	(B)	(C)	
1	Tax Register	Tax Register	Tax Register	Efficient and reliable for most pur-poses. Will include all enterprises; may or may not identify establishments. Will identify new and closed/dormant enterprises.
2	Census of Establishments	Census of Establishments	Census of Establishments	Census captures, most but not all, establishments. Too costly for annual use.
3	Single admin list	Single admin list	Single admin list	May be effective for entries but not for exits if closed units need not report exit.
4	Multiple admin lists	Multiple admin lists	Multiple admin lists	Requires matching the sources against the Business Register and each other, with risk of matching error. Will miss some entries. Exit data may not be available.
5	Single admin list	Single admin list	Annual survey or Business Register survey	Has advantages of (3) and may be effective for exits if the annual survey coverage is complete.
6	Multiple admin lists	Multiple admin lists	Annual survey or Business Register survey	Captures many, but not all, entries and may be effective for exits if survey coverage complete
7	Economic census	Single or multi-ple admin lists	Annual survey or Business Register survey	If matching of census and admin source is reliable, the two sources will complement each other for entries resulting in a more complete register.
8	Single or multiple admin lists	Census of Establishments	Annual survey or Business Register survey	

A less satisfactory solution than using the tax register or a single administrative list is to use a system that relies on multiple administrative lists. Such a system requires extensive matching to avoid duplication and much use of judgment at each step of the process. Complete data for exits are normally not available from the multiple sources (Case 4) with the result that the statistics office is obliged to document exits by other means (Case 6). The process is costly in management terms and there is a risk of failure should the attention of management falter. In many developing countries, however, there is no other choice in view of the lack of access to tax registers or to a single non-tax administrative source with adequate coverage.

A census of establishments can provide an initial list for a register but is not suitable for annual updating (Case 2) for simple reasons of cost. In addition, experience has shown that a census often misses a significant number of establishments, including recognisable ones.

Census coverage will be more complete if enumerators carry to the field a list from the existing register for their enumeration district instead of undertaking the census with a "blank slate". Finally, experience has shown that the combination of an occasional census and annual updating from administrative lists (Cases 7 and 8) can yield a more complete register than either the census alone or the administrative lists alone. This was demonstrated most recently in a UNIDO-supported project for updating the industrial register in Sri Lanka, as described in Annex 2.

8.0 Data structure of the Business Register

The data structure for Business Registers is usually limited to a few main topics:

- Coordinates of the unit, including contact person or persons

- Date (month and year is usually sufficient) of starting and ceasing commercial production. Also, date when the unit was added to the register and external data source by which unit was discovered (e.g. the VAT register, the Department of Health and Safety, etc.).

- Employment (or persons engaged) and any other indicator of size. A time series for the size indicator may be provided. The recommended employment concept is the number of persons engaged.

- Activity code and possibly multiple codes if secondary activities are shown or if a code under a previous classification is shown

- Activity status code, showing whether the establishment is active, closed, temporarily closed, out of scope, duplicated, etc.

- Current identification number, together with any previous identification numbers, such as for a previous register or for a census list

The coordinates of the unit may take up a large number of data fields:

- Addresses may be parsed into several fields; in Sri Lanka, they are parsed into eight fields. This action has two functions: to standardise the format and to facilitate matching.

- Similarly, phone numbers may be parsed into an area code and a local number. Names may be parsed into a name and a suffix or prefix indicating legal status.

- A geo-code for the location of each unit. In most countries, a system of geo-codes is available down to the lowest administrative level. In principle, it should be possible

to assign codes for statistical subdivisions (such as enumeration areas) of the lowest administrative level; in practice, however, this is rarely done for establishments. In either case, the consistent implementation of such a system cannot be taken for granted. If the system involves the lowest administrative level, its accuracy will depend on good knowledge on the part of establishment managers and/or enumerators from the statistical agency concerning the location of each unit. If the system involves statistical subdivisions, it will require very accurate and up-to-date maps showing streets and subdivision boundaries in urban areas.

- Additionally, for countries where finding an establishment cannot be taken for granted due to inadequate mapping or inadequate street addresses, Global Positioning System (GPS) coordinates would provide a reliable "address".

- For establishment registers, separate fields are needed for the name of the central office, address and phone numbers. Each central office (including those for single factories) must be treated as a separate establishment.

- For hierarchical registers, showing both establishments and enterprises and cross-links between them, as proposed in the 2008 Recommendations, the cross-references will also constitute data fields.

Finally, some data in support of the **Annual Survey of Industry** are also required:

- An indicator of whether the reports for the establishment are to be provided by the establishment itself or by a central office

- An indicator of which establishments are in the ASI sample for each year, if sampling is used

- A time series for response behavior to the ASI (as well as for other industrial surveys)

- Process variables for the ASI as described in the next section

9.0 The Updating Process

In many developing countries, there is no effective Business Register but there is an industrial register, for which a separate, well-defined register updating operation may or may not exist. Even if there is no updating operation, there is still an ASI and updating of the industrial register may be treated as an adjunct to the ASI. Field enumerators may be told to report any new establishments that come to their attention by asking them to fill out an ASI form. Enumerators may be asked to look out for new establishments as they

make their rounds or to inquire about nearby establishments (in a procedure sometimes called "snowballing", often used for sampling from social networks). Unfortunately, this casual approach has not been found to be very effective, for several reasons:

- Although enumerators may be exhorted to deliver annual survey questionnaires to all new/missed establishments, management lacks tools for checking enumerator compliance. Enumerators may deliberately fail to report large, well-known factories, and can do so without fear of reprimand. In the absence of documentation for the process of searching for new establishments, there is no way of knowing whether or not the enumerator has followed instructions. The enumerator himself has no way of knowing when he has "finished" the job. The system relies excessively on the motivations of enumerators, which can be expected to vary from place to place and time to time.

- If there is no channel for reporting the existence of new non-respondents, the only procedure for reporting a new establishment is to request it to fill out an annual survey questionnaire. If the establishment is asked to fill out the questionnaire but fails to do so, it remains invisible to the statistical system.

- Enumerators may be evaluated based on the percentage of their "target" (calculated as their share of the existing register) for which they delivered completed ASI questionnaires. The number of new establishments reported may not be used as an evaluation criterion. In this case, they may be unenthusiastic about reporting new establishments, which are often uncooperative, because they anticipate that they will have to return over and over again, at their own expense, before obtaining a completed questionnaire.

- Enumerators may receive no payment for searching for or discovering new establishments, although they often get piece-rates for other field work and these include reimbursement of both time and transport costs.

It is a definite advantage for a statistical agency to replace a casual system, such as that described above, with a clear procedure based on administrative lists even if these lists are multiple and incomplete. Use of such administrative data will give management more control over the process, as well as provide a basis for a more complete register.

10.0 The Updating System

An integrated application for register updating can greatly facilitate the process, particularly for the management of various files including the register itself and the external sources. For all but the smallest economies (ones with no more than a few hundred units),

the system should function on a Local Area Network. Systems built with a Structured Query Language (SQL) Server and .Net are internet friendly as well. For a system based on multiple administrative sources, seven basic modules are needed to support the following functions:

1. Importing of external data sources

2. Parsing and editing of data from the external sources

3. Matching of external sources against each other and the core register

4. Formation of the set of unduplicated entries from external sources that do not match to the register and a facility for prioritisation within the set to select subsets for field and phone checking

5. Entering field data for candidates from the field and phone checks and copying the "successful candidates" to the register after first checking for duplicates

6. Managing the register, including editing of records, *ad hoc* addition of records, and reports on the composition of the register

7. Managing the Annual Survey of Industry, as discussed in section 13

11.0 Business Demography

A well-ordered Business Register, one that is updated in a consistent way from year to year, can serve as the basis for reliable periodic statistics on business demography – entries and exits – in terms of number of units and their employment levels, as well as for employment size distributions by activity and district. Statistics for business demography provide useful indicators of business conditions by activity and district. Conversely, however, irregularities in time series for such tabulations can provide a useful indicator of flaws in the updating process.

In the context of business demography, the Statistics Office must take care to respect policies that will enhance the accuracy of statistics for business demography and minimise the apparent "churning" of business units, that is, the occurrence of fictitious exits and en-tries. The guidelines issued by the European Communities (see footnote 1, page 1) deal systematically with "continuity rules" for the issues mentioned here:

- When closures occur, the month and year of closure should be entered into the Business Register, so that closure is ascribed to the actual time of closure, not to the time that closure was reported. Such data will also be useful for imputing activity for that part of a year when the establishment was still operating.

- Care must be taken in deciding how to assign identification numbers to units, bearing in mind that any change in an ID number involves an implicit exit and entry. If an identification number is assigned to establishments based on the International Standard Industrial Classification (ISIC) and the district and sub-district, it is not recommended to change the ID number in the event of a change in the major product or in the location. This is particularly important because changes in the current ISIC or the sub-district code may sometimes be required due to errors in the initial assignment of the code.[6] For this reason, in the event of a change in code, the National Statistics Office in Indonesia adopted a policy around 1990 of distinguishing the current ISIC code from the original one in the industrial directory. The original ISIC was used as part of the establishment identification number (EIN) but the EIN did not change if the current ISIC changed.

- A clear Business Register policy is needed when a business physically relocates. In particular, should moves within the same district or sub-district be treated as an implicit exit and entry, or should the Business Register continue to use the same ID number after the move has taken place? The EU guidelines differentiate between moves over short and long distances.[7]

- Similarly, what about changes in ownership? Again, should the Business Register continue to use the same ID number after the ownership has changed? In these circumstances, there are two possible scenarios – one in which the major product remains the same, the other in which the major product changes under the new owner.

12.0 Hierarchy and Profiling

In line with the UN Recommendations, the Business Register needs to include both enterprises and establishments, with cross links between them. The shift towards increased interest in enterprise data is supported by the move toward increased reliance on administrative data, which may provide a good source for enterprise data. This contrasts with data from a census of establishments, which does not provide full cross-links to enterprises if conducted on the basis of door-to-door canvassing, although a well-designed census should provide comprehensive data on central offices as well.

---

[6] European Communities (2003), Chapter 18 "The Treatment of Errors" in *Eurostat Business Register: Recommendations Manual*.

[7] "It stands to reason to put much weight on the criterion of continuity of location, but this cannot be an absolute condition, because a move over a short distance of a local unit should be possible without loss of identity. If the same activity is continued with the same employment at a short distance from the old location, the move generally does not interrupt the local or regional function of the local unit."

In many advanced countries, the statistics office conducts a profiling survey from time to time during which enterprises are asked to provide data for their various establishments. Developing countries may wish to attempt a similar approach to register updating via enterprises since it would be difficult to collect consistent data for establishments which are subordinate to a single enterprise by means of establishment enquiries alone.

The guidelines for ISIC 3.1 provide extensive instructions on how to classify establishments and enterprises, including establishments with more than one activity[8]. According to the guidelines, "an establishment is an enterprise or part of an enterprise, which is situated in a single location, and in which only a single (non-ancillary) productive activity is carried out or in which the principal productive activity accounts for most of the value added." Under this concept, secondary products should not account for an important share of total production. Since, however, many establishments produce a variety of products, rigorous application of this concept would require intrusive profiling to split many "local units" (factories) into numerous notional "establishments" with their own outputs and inputs. The guidelines for ISIC 3.1 recognise that each country must decide for itself how far to proceed in this direction. In practice, some statistics offices are more willing than others to invest effort in disaggregating a local unit (such as a factory) into one-activity establishments.

- Some countries recoil from the burden on statistics offices and businesses of attempting to break down local units into more than one establishment, each with a different main activity and each activity with separately estimated inputs. In the USA and Canada, the tendency is for statistics offices to accept business practices. The US Census Bureau merely says that an establishment is "a business or industrial unit at a single location that distributes goods or performs services". Statistics Canada says that: "The Establishment is the level at which the accounting data required to measure production is available (principal inputs, revenues, salaries and wages)." Thus, neither agency attempts to parse the establishment into separate virtual units based on main activity.

- In contrast, European statisticians tend to avoid the term "establishment". They prefer the term "local unit" to describe what the US Census Bureau treats as an "establishment". The Europeans are more inclined than their North American counterparts to parse enterprises and "local units" into "kind of activity units" and "kind of activity local units". This, of course, requires the cooperation of businesses in conforming their reporting practices to statistical specifications.

---

[8] "International Standard Industrial Classification of All Activities", ISIC Rev 3.1, submitted to the United Nations Statistical Commission, 5-8 March 2002.

In practical terms, the difference between the two approaches to profiling mainly impacts on the treatment of secondary activities. In the North American approach, secondary activities may on occasion be reported separately (for example, on census forms) but inputs will never be reported separately for the various major activities of an establishment. Under the European approach, inputs and outputs would be reported separately on all occasions. The issue for developing countries is whether the convenience of separate reporting of outputs and inputs is worth the extra effort for both businesses and statistics offices of carrying out such profiling.

Annex 1: A System for Updating the Sri Lankan Industrial Register

The manufacturing industry in Sri Lanka was estimated in 2006 to account for about 18 per cent of Gross Domestic Product (GDP). A census of industry was carried out in 1983 and again in 2003/2004. In the intervening decades, however, register updating was partial and sporadic. In 1999, the United Nations Industrial Development Organization (UNIDO) began to provide advisory services on industrial statistics to the Department of Census and Statistics (DCS), which maintains the industrial register and conducts the Annual Survey of Industry, as well as to three other government agencies involved in industrial statistics (the Ministry of Industrial Development (MID), the Board of Investment (BoI), and the Central Bank of Sri Lanka. This initial work led to a UNIDO-supported project for updating the industrial register in two phases.

- In 2002, a project was launched with the aim of updating the register for the Western Province, where most medium and large industry is located. The project lasted 15 months and led to the discovery of some 700 establishments that had been missed by the core register and to the creation of a prototype system for computer-assisted register updating in Dbase and Visual Basic.

- In 2005, a second phase of the project started with the aim of updating the register countrywide. This 24 month project involved the creation of a more complete and reliable system for computer-assisted register updating, in SQL Server and .Net (dot net).

The experience presented below is taken from a paper presented at the third International Conference on Establishment Surveys (ICES). It is provided for the benefit of statistics offices that wish to begin updating their registers using a system based on multiple administrative sources.

A1.1 Core Register

The creation of a core register was carried out as follows:

- For phase one, a core register of 1,451 establishments in the Western Province already existed and was enhanced with annual ASI data for employment and response behaviour for the 12 previous years.

- For phase two, a core register had to be created in the aftermath of the 2003/2004 census. This was done by matching data from the pre-census register with census data and combining the unmatched records. The combined register comprised 5,235 establishments.

The core register data provide about 100 variables for each record, including:

- Identification numbers, including ID numbers for the 2003/2004 census of industry and the old core register

- Activity status, month and year of starting commercial production, month and year of closure

- ISIC 3 codes and main product

- Twenty variables for the names and addresses of the establishment and its head office, plus seven geo-codes. Addresses have been parsed into eight parts, names into two parts; these breakouts were designed to facilitate matching. This group of variables also includes geo-codes

- Another 20 variables for phone and fax numbers, which are each parsed into two parts

- About 26 variables for historical data for employment and ASI response status

- Dates for most recent updates of employment and activity status

- Names of owner and contact persons

- Data on annual kWh usage, based on electricity records for matched establishments

A1.2 External sources

For discovering new and missed establishments, the Department of Census and Statistics used four external data sources, with strengths and limitations as follows:

- The Board of Investment was the most effective source. The agency, which provides tax breaks and other concessions for investors, has a register of approved "projects" (separate establishments, for the most part) and knows which ones are still active. Addresses and phone numbers are up-to-date. Among BOI candidates, 66 per cent were successful in phase 1 and 64 per cent in phase 2.

- The Ministry of Industrial Development provides data on establishments registered with it but, unfortunately, many records are outdated, especially for activity status and telephone numbers. The source was effective in phase one; 47 per cent of candidates were successful. In phase two, the source was much less effective, with only 33 per cent successful on the field checks and 16 per cent on the phone checks.

- The Ceylon Electricity Board provides data on industrial customers. The data are up-to-date for activity status but mostly lack phone numbers. Establishment names are unreliable, as connections are often registered in the owner's name. In phase 1, 45 per cent of candidates were successful. In phase 2, 30 per cent were successful on the field checks, and 32 per cent on the phone checks. The field and phone checks showed many CEB listings to be duplicated with the core register but this fact had been missed during matching due to inconsistent establishment names.

- The Employees' Provident Fund provides data on establishments covered by the government-mandated social security system. The data included the number of employees but no phone number and an often outdated address. There were activity codes based on an *ad hoc* coding system. The codes had been added to the database long after registration by a private company to which the job was outsourced. This company lacked sufficient data for an accurate coding job with the result that the codes were unreliable. The source was found least effective in phase one, with only 17 per cent of candidates proving successful in field checks. Most unsuccessful candidates were either closed or out of scope (non-industry). The source was not used again.

A1.3 Matching

The external sources provided data in electronic format. After importing into the DCS system, parsing and editing, the lists from external sources were matched against the DCS core register. Matching was computer-assisted. The programme (in phase two) ranked possible matches using a matching likelihood index (MLI), as proposed by Weeks (2007). Operators reviewed blocked cases and those with a high MLI and classified them as either a match, non-match or pending. Pending cases were printed out for further analysis, often based on information to be elicited in phone calls. Matching during phase two began with a set of 7,308 records from external sources and concluded with 4,913 records from external

sources that were unduplicated in either the core register or in another external source and that could therefore be considered candidates for addition to the register.

A1.4   Prioritisation of candidates

Experience has shown that candidates from administrative sources may not qualify for the register for various reasons – for example, they may never have started the business, have closed, or be out of scope. It is therefore prudent to field check the candidates before adding them to the register. As funds were insufficient for field checking all 4,913 candidates, it was necessary to pick and choose. The candidates were, accordingly, divided into three priority groups, based on such considerations as quality of the source (BoI was top rated), size of employment or power usage, and registration year (with preference going to the most recent years).

- High priority candidates (883) were designated for field visits

- Medium priority candidates (705) were designated for phone checks

- The remaining low-priority candidates (3,328) were not checked because of lack of resources. The alternative would have been to check a sample of them and use the results as a supplementary frame for the register.

A1.5   Field and phone checks

The field and phone checks led to the discovery of 699 establishments, equal to 44 per cent of candidates checked. These included 282 establishments with more than 100 workers; total employment was 134,000, with an average of 192 workers per discovered establishment. Perhaps the most surprising finding came from the age distribution of discoveries. Half of these started commercial production before 2001, while only 29 per cent started in 2003 or after. This indicated that many of the discoveries were probably in-scope at the time of the census.

In a mature system for discovering new establishments, the backlog of undiscovered establishments should be small. This state could be reached in about a year given sufficient funding to check all "candidates" for addition to the register. In practice, however, it may take a few years to reach this state, given budgetary limits that would constrain the number of candidates checked each year. Accordingly, when discoveries for the current year are tabulated for a mature system, a large share should have begun commercial production recently. The finding that most discoveries for Sri Lanka are quite old suggests that the backlog of undiscovered establishments may still be quite large.

A1.6 Closures

For the Annual Survey of Industry (reference year 2005), the Department of Census and Statistics as usual mailed out questionnaires to all establishments on the register and followed up with phone calls and site visits for non-respondents. A null response form was used to document the reasons for non-response for 2,367 non-respondents and 271 cases that had either closed or gone out of scope. The closure rate, only 5.5 per cent, is suspiciously low, especially when compared with the finding of the call centre (see below) that about 13 per cent of establishments contacted by telephone were "dormant".

The null response form is supposed to be filled out during a field visit; however, there are grounds for concern that some of the forms were filled out without a field visit, in which case there is no positive, concrete evidence that the establishment was still in operation. If true, this would mean that many closures or "out of scope" cases were simply not observed.

A1.7 The register

The updating process for reference year 2005 began with a core register of 5,235 establishments. To this were added 699 discoveries, while 271 establishments were closed or went out of scope, leaving a net increase of 428. The updated register thus included 5,663 industrial establishments with 20 workers or more.

There was discussion of the possibility of publishing a directory of industrial establishments, which would both provide a public service and draw favourable public attention to the DCS' success in updating its register. However, the Department is not yet in a position to do this in view of the legal guarantee of confidentiality for all respondents, including industrial establishments.

A1.8 Low response rates

Low response rates have been a chronic problem for the ASI in Sri Lanka, with only 42 per cent of active sample establishments responding for reference year 2005. In an effort to improve the response rate, DCS established an ASI call centre with help from the UNIDO project. About 3,000 calls were placed, with most establishments receiving only one call. Most of those called said they had never received the questionnaire that DCS had mailed out; in such cases, the questionnaire was resent. The following response patterns to calls was observed: 13 per cent refused to have a dialogue or to respond to the survey, 20 per cent responded with combative questions, 30 per cent agreed to respond but sounded

negative, while the remaining 37 per cent agreed to respond. Given the prevalence of negative attitudes, it was concluded that DCS would need to place an average of five to seven calls per establishment in order to have a substantial impact on the response rates.

A1.9 Application

During phase two, a new application was built by a private software company, based on SQL Server and .Net. The new system functions on a LAN and is internet-friendly. There are six basic modules that support the following functions:

1. Importing of three external data sources

2. Parsing and editing of data from the external sources

3. Matching of external sources against each other and the DCS core register

4. Formation of the set of unduplicated entries from external sources and prioritisation within the set to selecting ones for field checking

5. Entering field data for candidates and copying the "successful candidates" to the register after first checking for duplicates

6. Managing the register, including editing of records, *ad hoc* addition of records, and reports on the composition of the register

The heart of the matching system is an algorithm based on analysis of bigrams – pairs of sequential letters in a character string. Similarity of an establishment name or street name – despite small spelling variations – leads to a high bigram score and thence to a high score on the matching likelihood index (MLI). The MLI formula was designed to give the benefit of the doubt to any possible match in order to ensure that the case would not be over-looked by operators. Thus, a high MLI could be reached in one of several ways: by having a near-matching name, or any matching phone number, or near-matching address (whether for the head office or establishment).

Operators review each case with a high MLI and decide whether it is a match, a non-match, or a pending case for further analysis. Pending cases are printed out for follow-up by telephone. As proposed by Weeks (2007), the programme was structured in such a way as to facilitate processing of cases with the highest MLIs first, in order to remove these cases from further consideration. The programme was also designed to enable supervisors to review staff decisions before they were implemented. Due to the user-friendly features of the application, DCS staff was able to complete the matching process in only a few weeks, much more quickly than during phase 1.

The entire updating system, as described above, is now fairly well established, having been built on the basis of the prototype from phase one, with detailed specifications for system elements that needed to be enhanced. During the phase two project, however, software problems frequently delayed work at DCS. Moreover, system maintenance and updating is now a major challenge for the Department, which lacks programmers experienced in .Net.

A1.10 Census of industry

In the time between the end of phase one and the start of phase two of the project, DCS carried out a census of industry. In preparing the census, it considered providing enumerators with lists of establishments from the old register but did not do this because of lack of time and lack of comprehensive and reliable geo-coding for the small areas where each establishment is located. The census yielded a list of establishments with 20 or more workers that was much larger than the pre-census register, indicating that many new establishments had been discovered. When, however, the UNIDO project supported the matching of the old register of 2,250 establishments against the census list of 4,920 establishments, the job turned out to be more laborious than anticipated as it involved hundreds of phone calls to identify matches and non-matches. In the end, DCS found that 375 of the establishments in the old register were still active in 2006 and had been missed by the census. Another 240 of the older establishments that were not matched to the census list could not be confirmed active; subsequently, spot checks at 10 of them showed that three were still active in 2006 and would have qualified for the census. This suggests that the census may have missed about 450 establishments from the old register (including roughly 75 from the unconfirmed group, based on the sample of 10 that was field checked), equal to 9 per cent of the census list and to about 20 per cent of the old register -- itself very incomplete. This estimate does not, of course, include establishments that were missed by both the old register and the census, many of which were discovered during phase two updating.

Given that the post census matching was carried out three years after the census listing, it was no longer possible to investigate why the census had missed these establishments. It was surmised that enumerators were often frustrated by the difficulty of gaining access to the establishments because security guards are instructed to keep visitors out. The large number of establishments missed underlines the importance of utilising data from the old register during the census field operations, since it is at this point that matching can be carried out on the spot.

Annex 2:   Screenshots from a Register Updating System in Sri Lanka

A2.1   The home page

A2.2 Establishments blocked as possible matches, with the Matching Likelihood Index

District COLOMBO ▼   Data Source BOI ▼   Likelihood LowestLikelihood ▼   Filter

1 2 3 4 5 6 7 8 9 10 11 12 13 14

Code	Establishment Name	H. MLI
3858/1998/17	TEA TECH SERVICES (PVT) LTD.	54
5544/2001/17	MILESTONE LANKA (PVT) LTD.	54
5473/2000/16	INTEGRATED SYSTEMS INTERNATIONAL (PVT) L	53
4438/1999/17	R AND R JEWELLERY (PVT) LTD.	53
2626/1995/17	YUAN FA LANKA AQUATIC PROD.CO.(PVT) LTD.	53
0023/1989/16	"PROF" COMPUTER MANUFACTURING (P) L	53
6116/2002/17	FIVESTARS APPARELS (PVT) LTD.	53
6681/2003/17	P.C.E. LANKA (PVT) LTD	52
6697/2003/17	ENVIRONMENTAL TECHNOLOGIES LANKA P. LTD.	52
7157/2003/16	NOMAD SHOWCASE SYSTEMS ASIA (PVT) LTD.	52
7315/2003/17	CHRISTY LANKA APPARELS (PVT) LTD.	52
0061/1993/16	DUTCH DAIRY INTL. (PVT) LTD.	52
2550/1995/17	BRITT LANGE INTL. (PVT) LTD.	52
3277/1997/16	PLANET GEM (PVT) LTD.	52
4803/1999/16	CHANG DAE INTERNATIONAL (PVT) LTD.	52
5288/2000/17	PREMATOR EAST (PVT) LTD.	51

A2.3 Blocked establishment details

5473/2000/16	INTEGRATED SYSTEMS INTERNATIONAL (PVT) L	53
4438/1999/17	R AND R JEWELLERY (PVT) LTD.	53

203.115.19.99:7070 - Match - Blocked Establishment Details - Microsoft Internet Explorer

_ UPDATING SYSTEM	CURRENT ASI YEAR 2005	Department of Census and S

- - -

SourceName - BOI
EstablishmentCode - 4438/1999/17
EstablishmentName - R AND R JEWELLERY (PVT) LTD.
DistrictName - COLOMBO

1

Target	Code	Estb. Name	MLI	Zoom	Matched
CORE	11026	A P A Jewellery	53	Zoom	✓
CORE	20210	LANKA JEWELLERY (Pvt.) Co. Ltd.	51	Zoom	✓
CORE	13632	Raja Jewellers	51	Zoom	✗
CORE	13648	Devi Jewellers	44	Zoom	✗
MIIP	5846	Wellawatha Nithyakalyani Jewellery	44	Zoom	✗
MIIP	993	Wit-Tasi Jewellery Manufacturers Pvt.Ltd	41	Zoom	✗
CORE	13654	Lalitha Jewellers	40	Zoom	✗

A2.4 Close-up of match options

http://203.115.19.99:7070 - Match - Matched Establishment Details - Microsoft Internet Explorer

NDUSTRIAL UPDATING SYSTEM	CURRENT ASI YEAR 2005	Department of Census and Statistics

X Close

Description	BOI	CEB	MIIP	DI2	CORE
SourceName	BOI	CEB	MIIP	DI2	CORE
Estb.ID	1353/1993/17				20210
Estb.Name	LANKA JEWELLERY (PVT) LTD.				LANKA JEWELLERY (Pvt.) Co. Ltd.
Telephone					4281003038
AssessmentNumber	20, 24				278
StreetName					Union Pl
CityName	PANADURA				Colombo 02
DSDName	Panadura				Panadura
DistrictName	KALUTARA				KALUTARA

Chapter 2: Concepts and Definitions of Data Items for an Industrial Survey

1.0 Introduction

The UN *International Recommendations for Industrial Statistics* 2008 (IRIS 2008) provide an extensive list of data items proposed for inclusion in Industrial Surveys, along with their definitions. This list has been used as a starting point for the following discussion but a number of changes have been made to facilitate implementation of Industrial Surveys.

The most important changes are:

- The use of terminology more closely aligned to the terminology of business accounts than the national accounts terminology of the International Recommendations. This ensures correspondence of the definitions with the data items actually included in the survey questionnaires

- A re-arrangement of the data items into groups that more closely follow the format of business accounts. This facilitates the completion of the questionnaire by the respondents and it also improves the possibilities for editing and verification of the data provided, both in the field and the office

- The addition of a number of questions, mostly qualitative, to obtain information on the general business climate; and

- Prioritisation of the data items to be included in surveys based on their use as well as the coverage of the surveys

As the perspective followed in these guidelines is implementation-oriented, a brief description of the survey methodology proposed is given first. This is followed by the presentation of the questionnaire design and coverage of data items for the different purposes of Industrial Surveys. Finally, the definitions of all the data items used in these questionnaires are given.

2.0   Survey methodology

The survey methodology proposed for industrial statistics is the FIRST methodology, described in the International Recommendations and originally proposed in *Strategies for Measuring Industrial Structure and Growth* (UN 1994).

One of the main characteristics of the FIRST methodology is the sub-division of the universe of industrial enterprises into two groups. The first group comprises all units included on a pre-determined set of lists. The companies included on those lists are generally large, easily identifiable and, because of the listing requirements, maintain complete sets of accounts. The enterprises in this group can easily be reached and mailed questionnaires are the most appropriate for this part. These questionnaires can include all items of information that are normally available in the accounts of a company. It is useful also to obtain copies of the audited accounts from the enterprises as these can be used for cross-checking while the notes to the accounts may provide additional information on data items that seem odd.

All other industrial enterprises are included in the second group. The number and location of the units included in this set is normally not known and these entities often do not maintain accounts or only have rudimentary financial information available. It is not practical to cover all units in this group and a sample survey is required. The most suitable format is area sampling where it is only necessary to make a full inventory of enterprises in the selected areas. The results of a recent economic census or, if not available, a population census, can be used to delineate the sample universe. The administration of questionnaires and the preparatory listing phase are conducted through interviews. The design of those questionnaires needs to allow for the fact that units without proper accounts cannot easily or accurately provide annual data. Moreover, information needs to be captured for shorter time periods and these are not necessarily the same for different data items. Provision therefore needs to be made for effective editing and verification of the data in the field.

The industrial enterprises in an area may also include units that are on the specified lists and therefore covered through the mailed survey. At the time of processing, such units need to be excluded from the data set in order to prevent double counting. While it is easy to notice large industrial units in any area, not all establishments of a large enterprise are big and easily identifiable. Consequently, the survey questionnaire used in the sample part should also include some questions that allow for the identification of units that are included in the list part.

Industrial Surveys, like most others, serve a variety of purposes. The questionnaire contains a number of standard sections but, depending on the uses of any particular survey, it may include some extra modules to supplement this basic information. A list of modules that may be included is given in Table 1.

Table 1: Questionnaire Sections

A. Basic module	B. Supplementary modules (variable)
1. Identification	11. Quantity and value of major materials used
2. Persons engaged, employment, wages and salaries	12. Quantity and value of production of major items
3. Sales, income from services and other revenue	13. Production capacity and its utilization
4. Expenditures, taxes and subsidies	14. Changes in sales and profitability
5. Changes in stocks and inventories	15. Major problems encountered in doing business
6. Profits and their appropriation	16. Research and development
7. Capital formation	17. Orders
8. Quantity and value of fuels used and produced	18. Traded goods, trade margin calculations
9. Quantity and value of electricity used and produced	
10. Quantity and value of water used and produced	

The contents of each module vary with the data requirements as well as their availability. For example, if it is planned to use the survey results in the construction of a Supply and Use Table (SUT), the level of detail of incomings and outgoings should be as close as possible to the detail of the product classification designed for the SUT. In other cases, it may be sufficient to broadly follow the level of detail maintained in the company accounts. The latter ensures a quicker response as much of the information can be copied directly from the accounts rather than needing to be freshly compiled from the detailed data.

The availability of information largely depends on whether or not an enterprise maintains accounts but, to a lesser extent, also depends on the legal reporting requirements in the country. The more elaborate those requirements are, the more information is routinely maintained by the corporate units. In countries with a system of value added taxes (VAT), there is generally a minimum threshold below which VAT is not charged and there are, therefore, no requirements to maintain even rudimentary accounts. It is also quite usual for a flat monthly or quarterly rate to be charged for businesses at the lower end of the taxable range within specific activities. Moreover, these units are not required to maintain accounts.

As mentioned above, for units that do not maintain accounts, the questionnaire should be limited to data items for which information is available, such as utility bills, or which can be easily recollected. In view of this, two basic questionnaires are proposed here, one for large units and the other for small units.

The distinction between the two is not based on the more traditional employment criteria but on whether or not accounts are maintained. The two main differences between the questionnaires for large and small units are the greater item detail in the large questionnaire than in the small one and the range of reference periods given in the small questionnaire to conform to the easiest recall period for the different data items.

Some additional issues are important in the design of the questionnaires.

First, items that are the same in both questionnaires should be assigned the same codes so that the integration of databases is simplified. Of course, those items that are different (such as the "other" groups) should be assigned unique, distinct codes.

Second, the list of items should include all components of income and expenditure listed in the company accounts so that information required for the integrated economic accounts can be derived from the same data set, in addition to that for the production account. The level of detail in the income and expenditure parts of the questionnaire is such that most items can be mapped to a single 4-digit ISIC activity even though these may not all be readily available from the published company accounts. On the other hand, it is quite common for company accounts to include the same items under both production and administrative costs and when this occurs the items should be consolidated before transfer to the questionnaire.

While the detail of the published company accounts varies widely between countries, and even between industries in the same country, the actual accounts tend to be maintained in much greater detail than is published. With the increased use of standardised accounting packages in many units, it is often also rather easy to obtain the extra detail required for the survey. In this respect, it is always much better to have estimates of sub-categories of an item given by the respondent rather than to have these imputed later on in the office, since the workers in an enterprise are much better informed about the operating details of their unit than the statistics office can ever be.

Third, the questionnaire structure should be such that it is easy to derive the profits (or loss) of the unit as the difference between income and expenditure. For the large units, this allows for the verification of the profit figures against those in the annual accounts, while the resulting profit figure in the small questionnaire can be confirmed with the respondent.

Fourth, some data items are listed in more than one place in the questionnaire. It is therefore necessary to ensure that the value for the same data item is the same in each of those places. In the questionnaire, all these different occurrences of the same information are cross-referenced. This helps those that complete the questionnaires and also provides a large number of edit checks for the IT staff to use in the data entry facility.

Fifth, the International Recommendations provide a large number of permutations of employment information. Most of these are not applicable to small units but, even for the large questionnaire, choices will have to be made. For the various employment categories (working proprietors, unpaid workers, employees, etc.), it is recommended that the basic information requested on numbers and wages and salaries (including various sub-components) should cover:

1) Male and female workers

2) Operatives and other workers

3) Full-time and part-time workers

4) Workers' skill levels   (not in the Recommendations but important for many countries); and

5) Local and foreign workers (not in the Recommendations but important for many countries)

The full complement of tables needed to obtain all the recommended employment and wage and salary details runs into tens of pages but, for smaller enterprises, most of the data cells will be empty. It is therefore important to establish the genuine need for various types of detail in order to conserve resources and to obtain a timely response. The large questionnaire contains a number of different data blocks to capture the various dimensions of employment information and these may be used as examples in individual countries with more emphasis on some dimensions of the information than on others.

In line with business accounting, the information sought in the questionnaire relates to sales, receipts, and incomes during the year. Expenditures are recorded in a similar way although in business accounts it is more usual to enter sales revenue than purchases of materials, in addition to other expenditures In national accounting, production and cost of production are the standard concepts. The difference between those concepts lies in the timing of the transactions (over the lifetime of an enterprise, production costs and sales revenue will be the same, as will be purchases and the cost of production) and the reconciliation is done (in Section 6 of the questionnaire) through adjustment with inventory levels (given in Section 5).

As the trading activities of industrial units are often carried out to support the main activity, rather than as a separate activity, they need not be profitable in themselves. However, in order to increase understanding of the operations of the enterprises, it is useful to evaluate the economics of trading in the case of industrial companies.

In general, industrial enquiries only cover those establishments which have started operations. This may be for practical reasons, such as the date of their registration, or it may be based on the understanding that the principal purpose of the survey is limited to the estimation of production and value added of the sector; whatever the reason, establishments which are not yet in production are quite frequently not covered in the survey. However, since a very large part of investment in plant and machinery is done well before production starts and the construction period often stretches over a number of years the omission of establishments that are under construction results in under-estimates of the level of investment in industry.

In view of this, and at the very least, the treatment of pre-production capital expenditures should be documented in the published survey results in order to prevent incorrect analysis of the survey information. For example, capital output ratios and various productivity measures will be skewed if pre-production investment is not included in the equation. A better solution, however, is to include establishments under construction in the survey and thus obtain a comprehensive set of data on capital formation.

3.0   Definition of individual data items – Large Questionnaire

Section 1. Identification

This section consists of three different parts. The first part, *Identification*, gives the particulars of the establishment, including name, address, contact details, as well as the establishment number assigned for the survey (either a survey-specific number or the code used in the business directory). In addition, the name and telephone number of the contact person are given here.

The second part describes the *activities* of the establishment. If there is only a single activity carried out by the establishment, this is by default its principal activity. Otherwise, the principal activity is defined as the activity that generates the highest income. All other (significant) activities are also described in this part. All activities are coded using the ISIC classification of industrial activities.

It should be noted that an establishment can change activity over time and the survey information on its activities should be used to update the information on it already included in the Business Register or list. This also ensures that grossing-up factors for non-response can be more accurately calculated by activity.

The third part of this section describes the establishment's *organisational status*. Various data items are included here.

First, information is requested on the establishment's legal status. While the questionnaire lists the most common types of organisation, each country has its own legal environment and consequently its own types of legal status. It is therefore important to modify this list, where necessary, to ensure that all the main legal categories of an establishment in the country in question are included and that the locally used names are given.

Second, the type of ownership is requested. This is largely a permutation of three main owners - government, local private, and foreign. Where different levels of government are active participants in the enterprise sector, these may be listed separately (e.g., central, regional and local).

Third, the name and address of a parent company, if any, needs to be included. Of course, it is also possible that an establishment has subsidiary establishments itself and is thus a parent company. However, since only a few industrial establishments have subsidiaries, this item is not usually included. Nonetheless, in those countries where this does occur, a specific question can be added.

The fourth data item is the year of commencement of operations. If a Business Register is maintained, this information is probably already available. However, this is not always the case and, consequently, it may still be useful to include the question in case a particular establishment is not on the register.

The fifth item requests the number of months of operation during the year, along with a question as to whether the establishment is seasonal or not. For seasonal units, the number of months normally operational per year is then assumed to be the number provided against the earlier question in the large questionnaire.

The sixth and final question on this page identifies whether or not the unit maintains accounts. For those who do, the dates of the accounting year are also requested. It is the answer to this question which determines whether the large or small questionnaire should be used. For units that do maintain accounts, the large questionnaire is applicable while for those not maintaining accounts, the small questionnaire is used.

Section 2. Employment, wages and salaries, hours worked

Employment

Employment information, and wages and salaries for those categories to which these apply, is asked for a number of different types of workers. The employment types listed in the questionnaire include:

- Working proprietor – the person owning and operating a household enterprise. A working proprietor does not receive regular wages or salaries

- Unpaid family or other workers – persons, generally related to the owner, who work without receiving regular payment (that is, without an agreed amount to be paid for work done), either in cash or kind

- Apprentices – students acquiring necessary skills through an on-the-job training programme. Apprentices may receive some compensation but they do not normally receive a regular salary

- Working director – director of an incorporated unit involved on a daily basis in the operation of the unit

- Permanent employees – workers on a long-term employment contract who receive regular payment for their labour, in cash and/or in kind

- Casual/seasonal employees – workers engaged on a daily or short-term basis who receive regular payment for their labour, in cash and/or in kind

- Leased workers – staff provided through an outside agency (employment agencies, sub-contractors, etc.) on a short-term or long-term basis, who are not on the payroll of the company but on that of the agency providing the staff

- Working proprietors and unpaid workers are not counted as employees

Various other groupings of workers can be defined, such as:

- Outworkers – workers engaged by the company who fit into any of the categories above (except leased workers) and who are not working on the company's premises. This category includes travelling salesmen, field maintenance workers and so on. Outworkers paid by subcontractors are not included; the amounts paid to subcontractors in respect of outworkers are treated as "cost of services purchased"

- Production workers (operatives) – workers directly engaged in the production process. This includes machine operators, manufacturing clerks, production management staff, etc.

- Other workers – workers not directly engaged in the production process. It includes all workers not counted as operatives

- Full time workers – workers who are, during their period of engagement, working the regular full working week as determined by law, contract or otherwise. This also

inclludes those seasonal or casual workers that work for the full working day or week during the period they are engaged

- Part time workers – workers that are engaged for less than the full working week; this applies to both permanent and temporary workers

- Skills levels of the staff – skilled, semi- skilled, unskilled, apprentice, etc.

- Workers by functional category – e.g. staff engaged in own-account construction work, workers in fabrication, processing and assembly, transportation or warehousing, repair and maintenance staff

For productivity analysis, it is important that the various categories of workers listed above (except the outworkers) include both the company's staff as well as the leased staff. Persons on short-term leave (sick leave, maternity leave, annual leave or vacation) and on strike are also to be included in the various counts. In the questionnaire, the persons engaged by the establishment are further sub-divided by gender (males and females) and by nationality (locals and foreigners).

Wages and salaries

The total remuneration of employees, in cash or in kind, is the amount payable by an enterprise to an employee in return for work done by the latter during the period. It is made up of two main components:

1) Wages and salaries in cash or in kind:

- Wages and salaries in cash, including allowances (housing, travel, etc.), bonuses, commissions and gratuities

- Wages, salaries and benefits in kind (e.g. meals and drinks, free accommodation, uniforms, use of company vehicles outside work, child care facilities, subsidies on interest on loans given to employees, value of stock options given)

- Social insurance contributions, health insurance, income tax and other deductions, payable by the employee but withheld by the employer for transfer to social security schemes, tax authorities, etc.

2) Social insurance contributions payable by employers.

Payments to directors of incorporated enterprises and members of shareholders' committees for attending meetings are also included under compensation of employees.

However, in the questionnaire, this item is only included under item 436 in section 4 (expenditures), as it is only available as a single data item without further breakdown.

Hours worked

The number of hours worked by all persons in employment (paid and self employed persons) is defined as the total number of hours actually spent on activities by employees, including waiting time, that contribute to the production of goods and services during the year. Based on the total number of hours worked, the number of full-time employment equivalents can be estimated. This number is derived by taking total hours worked in a unit divided by the average number of hours worked by a full-time employee.

Section 3. Income

In this section, all items that represent current income for the company are listed so that comparison with the company accounts can be easily made. While not all information is used for the production accounts, much of the extra information is valuable for the compilation of the integrated economic accounts of the nation (commonly referred to as the institutional sector accounts). To obtain the same information by other means would either require special surveys or the analysis of published annual accounts. Both of these approaches are much more costly and time consuming than the inclusion of these items in this questionnaire. However, the main reason for including all these items is that it makes editing simple, as the profits derived in the questionnaire (in section 6) should be equal to those stated in the company accounts. This not only applies to units maintaining an independent set of accounts but often also to individual establishments that are treated as "profit centres", even though the overhead costs may be assigned arbitrarily by the head office.

In line with standard business practices, in countries where the value-added tax system is used, the value-added tax received from clients, as well as that paid to suppliers, is excluded from the company accounts.

The following items of information are separately requested in this section:

301	Value of sales from own production (finished products and by-products) ** ** item 301 should equal item 1299(d) in section 12
302	Value of sales of semi-finished goods
303	Receipt from sale of used goods by the establishment *(items such as sacks, drums etc.)*
304	Receipt for sale of waste and scrap products
305	Receipts for contract work done (on goods not owned by the company)
306	Receipts for repair and maintenance work done
307	Value of own produced new construction, improvement or renovation made on fixed assets for own use ** Item 307 should equal item 799(c) in Section 7
308	Value of sales of goods bought and resold in the same condition as purchased *(traded goods)* ** Item 308 should equal item 1801(a) in Section 18
309	Value of sales of water ** Item 309 should equal item 1004(b) in Section 10
310	Value of sales of electricity ** Item 310 should equal item 903(b) in Section 9
311	Value of sales of waste water treatment services
312	Receipts for transport services rendered *(including charges for delivery of goods sold, when billed separately)*
313	Income from commissions, etc.
314	Income from rental of buildings
315	Income from rental or lease of machinery and equipment
316	Income from the sale or licensing of R & D products ** Item 316 should equal item 1669 in Section 16
317	Income from rental of land
318	Interest received
319	Dividends received
320, 321	Other earnings on investments *(please specify)*
322	Subsidies received
323	Export rebates received
324	Gains on sales of fixed assets
325	Receipts from insurance claims
326	Exchange gains
327 to 329	Other income *(please specify)*
399	Total income (sum of items number 301 to 329) *(please transfer item 399 to item 602 in Section 6)*

The value of sales is the amount billed to the customer. It should be the actual value received or receivable, after all discounts, commissions, and so forth have been deducted. The valuation can include transport costs and should also include all duties and taxes

imposed on products when they leave the factory, with the exception of the value-added tax invoiced by the producer to the client, where the value-added tax system is applicable.

Although the listing of types of income covers most items common for industrial units, it provides additional space for free entry of some items (against lines 320, 321 and 327 to 329) not specifically listed. Each item also given elsewhere in the questionnaire is cross-referenced to provide guidance to the respondent. Editing, both manually in the field and computerised at the time of data entry, can use the same cross-checks.

The valuation of transactions between establishments of the same enterprise should be done as if sold. In practice, this may not always be possible and the production cost has to be used for such transactions. Production cost is equal to the sum of material and service costs, compensation of employees, other taxes on production, depreciation of the fixed assets used in production, and, sometimes, an imputed margin for overhead costs and profits. In general, this is also the valuation used for finished products entering the inventories.

Details on the quantity and value of sales of individual, important products are recorded in Section 12. The production capacity and its utilisation during the year by individual products/product lines is specified in Section 13.

Section 4. Expenditures

In line with the procedures described under Section 3, expenditures include all current expenditures (excluding VAT) incurred during the year. This includes the following items, cross-referenced to other sections of the questionnaire where applicable:

401	Payments made for raw materials, chemicals, packing materials and spare parts **Item 401 is equal to item 1199(d) in section 11
402	Payments made to other firms and individuals for work performed on sub-contracts
403	Payments made for repair and maintenance work
404	Payment made for fuel purchased *(not for household use)* ** Item 404 is equal to item 819(d) in Section 8
405	Payment made for electricity purchased *(not for household use)* ** Item 405 is equal to item 901(b) in Section 9
406	Payment made for water purchased *(not for household use)* ** Item 406 is equal to item 1001(b) in Section 9
407	Purchase of goods for resale *(traded goods)* ** Item 407 is equal to item 1811(a) in Section 18
408	Commission Paid
409	Payment for rental of warehouses, industrial space, etc.
410	Payments for rental of office, shops & other premises
411	Payments for rental or lease of machinery and equipment
412	Payments for installation charges
413	Transportation charges paid
414	Payments made for telephone, telegraphs & telex
415	Payments made for internet access
416	Payments made for postage, couriers, etc.
417	Payments made for printing and stationery
418	Purchases of books and journals
419	Payments of travelling costs for business purposes
420	Payment of insurance premiums
421	Bank charges *(excluding bank interest charges)*
422	Payments made for advertising and promotion *(excluding printing)*
423	Payment of management fees
424	Payments made for legal expenses
425	Payments for auditing and accountants' fees
426	Payments for other professional services
427	Payments made for license/registration fees
428	Payment of royalties, copyrights, patents, etc.
429	Payments for educational and training
430	Cost of treatment and disposal of waste products
431	Depreciation charges for the year on fixed assets ** item 431 should equal item 799(h) in section 7
432	Interest paid

433	Payments for rent of land
434	Payments of taxes on production *(e.g. Sales Tax, Excise Tax, Hotel Tax, Mineral Duty, Timber Levy etc.)*
435	Payments made for wages and salaries ** item 435 should equal item 299(h) in section 2
436	Directors' remuneration, attendance fees and allowances
437	Provision for bad or doubtful Debts
438	Loss from sales of fixed assets
439	Loss from investments and stocks
440	Exchange losses
441	Cash donations, grants & gifts
442 to 444	Extraordinary items *(please specify)* *(this should include all items in the annual accounts not specified above)*
499	Total expenditure (sum of items number 401 to 444) *(please transfer item 499 to item 604 in Section 6)*

The valuation of goods purchased should be in the purchasers' prices – that is, the delivered value at the establishment, including the purchase price, transport charges either invoiced by the producer or by other organisations, the cost of insurance, the value of packaging materials charged for, all taxes and duties on the goods (excluding the deductible VAT, where applicable). Discounts and rebates allowed to the purchaser should be deducted, as well as the value of packaging materials returned to the suppliers.

Quantities and values of important items used in production are specified in Section 11. The details on quantities and values of various fuels purchased are given in Section 8. Details on the production and consumption of electricity are given in Section 9 and those for water and waste treatment are listed in Section 10.

Section 5. Inventories and stocks

The value of all inventories owned by the establishment, either at the establishment or elsewhere, should be included. As the economic ownership rests with the unit, inventories held at ancillary units, in bonded stores, public warehouses or overseas, on consignment, in transit, and materials being manufactured, processed or assembled on commission by others should all be included. However, materials owned by others but held by the establishment for processing should be excluded. Separately listed are inventories of:

501	Raw Materials
502	Chemicals and dyes
503	Packing material
504	Spare parts
505	Fuels and lubricants
506	Work in progress
507	Finished goods
508	Goods purchased for resale
509	Any other stocks and inventories
599	Total *please transfer item 599(a) to item 605(b) in Section 6.* *please transfer item 599(b) to item 601(b) in Section 6.*

Valuation of inventories should, as far as possible, be consistent with the valuations of finished products sold and materials and supplies purchased. Inventories of materials, fuels and supplies acquired from others should be valued at purchasers' prices. Work- in-progress and inventories of finished goods should be valued at equivalent basic prices (market prices excluding taxes on products, transport costs and trade margins) or at production costs if equivalent basic prices are not available.

Section 6. Profits

The information collected in Sections 3, 4 and 5 can be reconciled with the annual accounts of the company. This is done in the framework given in this section, the principal purpose being to ensure the consistency of the various sections of the questionnaire and between the questionnaire responses and the annual report of the company.

601	Value to stocks at end of accounting year *(copy item 599(a) in Section 5)*
602	Total income *(copy item 399 in Section 3)*
603	Sub-Total (601 + 602)
604	Total expenditure *(copy item 499 in Section 4)*
605	Value of stocks at beginning of accounting year *(copy item 599(b) in Section 5)*
606	Sub-Total (604 + 605)
607	Profit before taxation** Sub-Total (603 – 606)
608	Unappropriated profits brought forward from previous year**
609	Sub-Total (607 + 608)
610	Profit retained in business **
611	Profit distributed as dividends, etc. **
611.1	of which: profits distributed to foreign shareholders
611.2	profits distributed to local shareholders

612	Sub-Total (610 + 611)
613	Unappropriated profits carried forward to following year** (equals 609 – 612)
	*** Items 607, 608, 610, 611, 613 should equal the same items reported in the annual accounts of the establishment.*

For establishments that do not have complete accounts, the information in items 608 to 613 is not relevant. For other establishments, the information is used in the compilation of the institutional sector accounts mentioned earlier.

Section 7. Fixed assets

All goods, except minor tools and equipment, expected to have a productive life of more than one year and intended for use by the establishment (land, mineral deposits, timber tracts, buildings, machinery, equipment and vehicles) are considered capital goods. Also included in this section are additions, modifications and improvements to existing fixed assets that extend their normal economic life or increase their productivity. Fixed assets may either be purchased or produced by the unit's own labour. In the latter case, the value of the asset is also given as an output in Section 3 (item 307). Also included are fixed assets purchased under financial lease, that is, where lease payments over the contract life will equal the asset's value.

The fixed assets include the following items:

701	Land
702	Land improvement
703	Residential buildings
704	Other buildings
705	Other structures, e.g. roads, bridges, mine shafts, etc.
706	Transport equipment**
707	Computers and related equipment**
708	Other plant and machinery**
709	Capital work in progress
710	Livestock for breeding, dairy, draught, etc.
711	Vineyards, orchards and other plantations of trees yielding repeat products
712	Mineral exploration
713	Computer software
714	Entertainment, literary or artistic originals
715, 716	Other intangible fixed assets *(specify)*
799	Total
	***Include machinery under financial lease (i.e. where lease payments over the contract life will equal its value)*

As information on the stocks of fixed assets, as well as changes in their composition, are important data items, for each of the asset types specified against items number 701 to 716, the following stock and transaction information is registered:

(a) Book value of assets at start of year

(b) Purchases of fixed assets

(c) Own account capital formation

(d) Revaluations of fixed assets

(e) Sales/disposals of fixed assets

(f) Fixed assets lost or destroyed

(g)= (b) + (c) +(d) – (e) – (f) Gross additions to fixed assets

(h) Depreciation charges for the year

(i) = (a) + (g) – (h)   Book value of assets at end of the year

(j) Depreciation rates used

Valuation of the assets and their transactions follows normal business practices. These are often different from the definitions and treatment used in the national accounts which should use all information available to derive their data. For example, in the business accounts, item (e) above would cover only the depreciated value of the asset, while any difference between this value and the actual sales value will be recorded as gain or loss on the sale of fixed assets, items 324 and 438 in Sections 3 and 4 respectively.

Similarly, in the national accounts, unsuccessful mineral exploration activities are capitalised. In the company accounts, however, they are treated as intermediate consumption and therefore not included in item 712 of fixed assets.

A third difference is the treatment of capital work in progress (mostly construction work and plantation development), which is generally included in the company accounts under fixed assets (item 709 above) but is treated as work in process under inventories (item 506) in the national accounts, when purchased and treated as fixed capital formation when done on own account.

Section 8. Fuels

In order to calculate energy consumption, it is necessary to collect the quantity of the individually important fuels purchased and the quantity of electricity purchased, generated and sold. In this section, the quantities and values of fuels are recorded while electricity is recorded in Section 9.

	Solid fuels
801	Fuel wood
802	Charcoal
803	Coal
804	Coke
	Liquid fuels
805	High speed diesel oil
806	Light diesel oil
807	Gasoline/petrol
808	Kerosene
809	Furnace oil
810	Engine oil
811	Liquefied gas (LPG / CNG)
	Gaseous fuels
812	Natural gas / factory (coking) gas
813 to 815	**Other fuels & lubricants** *(please specify)*
819	**Total fuels purchased** (801 to 815)
	Minus changes in stocks of:
821	Fuels and lubricants
	Plus fuels recovered from waste
822	Methane recovered from waste products *(this should equal item 1052 in section 10)*
823	Heat recovered from incineration of waste *(this should equal item 1054 in section 10)*
829	**Total fuels consumed** (819 to 823) *of which*
899	**Total purchased fuels used** (819 and 821)
	Note: item 821(c), if not available, can be estimated using the ratio of items 819(c) / 819(d)

The quantities of individual fuels consumed are recorded in the unit of measurement commonly used for that item in the country. For aggregation of the quantities of fuels used, it is necessary to convert them to a common unit. The internationally recommended unit for this is the terajoule. Conversion for the various fuels into this unit can take place in the office but it may be useful to provide the information also to the respondents. The questionnaire therefore includes the conversion factors for all fuels listed and the calculations can thus be done in the field.

In order to assess the use of renewable energy resources, Section 8a enumerates the use of fuels from these sources, as listed below. Three different sources of renewable fuels have been listed, namely the traditional fuel wood and charcoal (although in this section they should be reported only if obtained from sustainable sources), synthetic liquid fuels and fuels and heat recovered from waste products.

Section 8a. Fuels:- of which from renewable sources

	Solid fuels
851	Fuel wood
852	Charcoal
	Liquid fuels
853	Bio diesel
854	Ethanol
855, 856	**Other fuels & lubricants** *(please specify)*
857	Methane recovered from waste products *(This should equal item 1052 in section 10)*
858	Heat recovered from incineration of waste *(This should equal item 1054 in section 10)*
859	**Total from renewable sources**
	Percentage from renewable sources *(item 859/item 829%)*

When all usage is expressed in terajoules, the share of fuels from renewable sources in the total can be calculated.

Section 9. Electricity

Electricity is used extensively in many production processes and for auxiliary functions of the establishments. Many units produce at least part of their own needs and the sale of surplus electricity has traditionally also been part of the activities of some industrial units.

901	Electricity purchased (or obtained from other establishments of the same enterprise)
902	Electricity generated
903	Electricity sold
999	Total Electricity used (901 + 902 - 903)
	Item 901(b) is the same as item 404 in Section 4.
	Item 903(b) is the same as item 310 in Section 3.

The generation of electricity from renewable resources is enumerated separately in section 9a as indicated below. As all electricity is measured in the same unit of quantity, a simple calculation shows the percentage generated from renewable resources.

Section 9a. Electricity: of which generated from renewable sources

951	Hydro
952	Solar
953	Wind
954	Biomass/bio-fuels
955, 956	Other renewables *(specify)*
959	Total electricity generated from renewable sources
	Percentage from renewable sources (item 959/item 999%)

In parallel with other manufacturing processes, electricity production is calculated on the basis of the supply available for sale. In captive units, this is the gross generation minus station use. For commercial electricity generators, the line losses are also deducted from the gross production so that the net supply available to customers is measured.

Section 10. Water and waste treatment

Water is an important input in many production processes and the large flows of wastewater generated are often major sources of pollution, especially in urbanised environments. Traditionally, the emphasis has been on water, which is sometimes scarce, and often costly, while wastewater is drained off without regard for the environment. With the increased attention to pollution abatement, information on both water and wastewater (and waste in general) has become more important. At the same time, the two flows, which were largely separated in the past, have developed more linkages. It is nowadays not uncommon, including in developing countries, to find enterprises that recycle and reuse all their water.

In this section of the questionnaire, the following information is collected for water:

1001	Water purchased
1002	Water produced / pumped up / collected / drawn in
1003	Water reused / recycled
1004	Water sold
1099	Total water used (1001 to 1003 – 1004)
	Item 1001(b) is the same as item 406 in Section 4.
	Item 1004(b) is the same as item 309 in Section 3.

In addition, the following information is collected on the processing of waste flows and the reuse of its products, if any:

Section 10a. Waste treatment

1051	Industrial wastewater treated (to local environmental norms)
1052	Recovery of gas (methane)/fermentation
1053	Composting
1054	Incineration of solid waste/heat recovery
1055	Treated water reused/recycled

Section 11. Raw material details

For the most important raw materials used in the production process, information on the quantity and cost of those individually important materials is collected. The information in this section is mainly used for the construction of supply and use tables. Since, in most countries, these tables are not often produced, these data do not always need to be collected.

In many countries, it is useful to supplement the quantities and values of the items purchased with information on the quantities and values of the same items directly imported by the establishment.

Section 12. Production details

For the most important items produced by the establishment, information on the quantity and value of those individually important products is collected. The information in this section is mainly used for the construction of supply and use tables and, since in most countries these are not often produced, these data do not always need to be collected.   In addition to information on the quantity and value of sales in total, it may also be useful to collect information on the direct exports of the same products by the establishment.

Section 13. Production capacity and its utilisation

For various types of analysis, information on the production capacity of the manufacturing industry, and its utilisation, is important. Depending on the type of industry and the

production processes used, a product line may produce a single type of product or a whole range of outputs. As far as possible, the production capacity should be reported for the major products (or groups of them) individually listed in Section 12. In some types of manufacturing, production capacity cannot properly be measured and, in such cases, the section may be omitted.

Section 14. Sales and profitability

In this section, the perceptions of the management or owner of the establishment on trends in sales and profitability are enumerated. For both characteristics, a comparison is made with the previous year and with the four previous years in order to measure longer term trends. The possible answers are on a five step scale:

(a) Much better

(b) Somewhat better

(c) About the same

(d) Somewhat worse

(e) Much worse

While this qualitative information is not very useful by itself, in aggregate, the responses give a good indication of past trends in these important characteristics in different industries.

Section 15. General business environment

In a second set of qualitative questions, the areas where major problems may be faced by the enterprises are enumerated. Of course, the situation in each country is different, and appropriate modifications to this list should be made before administering the questionnaire.

1501	Availability of affordable finance / bank loans
1502	Availability of tools and machinery
1503	Availability of working space
1504	Availability of skilled/technical labour
1505	Availability of unskilled labour
1506	Availability of electricity
1507	Availability of water

1508	Availability of input materials
1509	Availability of transport facilities
1510	Government/regulatory environment
1511	Marketing problems
1512 to 1516	Other problems *(please specify)*

For each of the potential problems, five options are given for an answer. In addition, the question is asked whether the situation has improved or deteriorated in comparison with the situation four years earlier.

(a) Very serious

(b) Serious problem

(c) Minor problem

(d) No problem

(e) Not Applicable

(f) Compared with four years ago, has the situation improved?

(g) Compared with four years ago, has the situation deteriorated?

Section 16. Research and development

Information on research and development (R&D) has not received much attention in Industrial Surveys in the past. This section looks at three different aspects of the R&D efforts in a company. First, in Section 16a, the number of staff involved in the five types of R&D work distinguished is entered. The layout is identical to that used in Section 2 but provides specific information on the R&D staff. The five activities listed are:

- Intellectual property products

- Mineral exploration and evaluation

- Computer software and databases

- Research and development

- Entertainment, literary and artistic originals

In the two following sub-sections, information on the value of R & D work that has been capitalised and the income from sales and licensing of R & D products is sought. In Section 16b, the amounts of work capitalised are given, also for the five categories. The amounts are transferred to the column "own-account production of fixed-assets" in Section 7, from where they are copied to the income section (item 307).

The value of sales or licensing of R&D products is summarised in Section 16c and transferred to income (item 316).

Persons employed in research and development, computer software and databases comprise all persons employed directly on research and development, computer software and databases, as well as those providing direct services, such as research and development managers, administrators and clerical staff. The personnel must be distinguished from personnel for a wide range of related activities. The following are therefore **excluded** from research and development personnel:

- Personnel employed in education and training

- Personnel employed in other scientific and technological activities (e.g. information services, testing and standardization, feasibility studies, etc.)

- Personnel employed in other industrial activities (e.g. industrial innovations n.e.c.)

- Personnel employed on administration and other indirect supporting activities

Section 17. Orders

In selected branches of industry, the information on new orders received and filled can be very useful in tracking the strength or weakness of investment and production in the near future. For each of the main products of those enterprises (as reported in Section 12), the development of the order status is followed using the following information:

(a) Commodity code

(b) Unit of measurement

(c) Unfilled orders at the beginning of the year

(d) New orders received during the year

(e) Orders filled during the year

(f) Orders cancelled during the year

(g) Unfilled orders at the end of the year =(c+d-e-f)

Section 18. Traded goods and trade margins

Many industrial units also engage in trading activities, often not by choice but in order to provide adequate support and service through the provision of spare parts and auxiliary items or to extend the range of products they can offer their customers. In order to have a better understanding of the economics of this part of their activities, this section is used to derive the gross trade margin.

1801	Sales of goods bought and resold in the same condition (traded goods) ** Item 1801(a) should equal item 308 in Section 3
1802	Goods withdrawn for own use (if any)
1803	Goods given to employees (if any)
1804	Closing stock at the end of the year ** Item 1804(a) should equal item 508(b) in Section 5
1809	Total use of goods purchased for resale (1801 to 1804)
1811	Purchase of goods for resale in the same condition (traded goods) ** Item 1811(a) should equal item 407 in Section 4
1812	Opening stock at the beginning of the year ** Item 1812(a) should equal item 508(a) in Section 5
1819	Total supply of goods purchased for resale (1811 and 1812)
1899	Gross trade margin (1809 – 1819)
1851	AVERAGE percentage margin added to purchase price (1899 / 1819)

4.0 Definition of individual data items – Small Questionnaire

The small questionnaire has the same overall structure as the large questionnaire but only includes a subset of the items, while some of the sections have been omitted altogether as they are not relevant for small units. Nonetheless, the draft included is rather elaborate and, especially when the focus of the survey is on household units, the questionnaire can be pared down further.

The main difference with the large questionnaire is that much of the information needs to be collected on the basis of recall by the respondent as no formal accounts are maintained by the establishments. In order to improve the quality of the recall, it is important to find the most appropriate reference period for a particular data item. It is important to note that the reference period varies for the different items.

A few examples illustrate this. If an enterprise pays annual registration fees, quarterly vehicle licence fees and monthly electricity bills and buys top-up cards for the mobile phone when the time is exhausted, the most appropriate reference periods for the first three items are year, quarter and month respectively, while further investigations are required to determine the best reference period for the phone cards. Yet even for the electricity bills, this is not straightforward as the most likely response will be that the monthly amount varies and is reported to be between x and y. In general, if it is too difficult to get the actuals from the bills, the mid-point of the range given by the respondent is a fair estimate.

Rather than providing the reference period to the respondent, it is often best to keep the initial question open and react to the response. The answer to the question "How much do you sell" may give the respondents' logical reference period, which may be completely different from what might be expected. A unit that ships its products on a daily basis may quote daily production but a unit that ships only when a truck-load is complete, or on an agreed delivery schedule of its main customers, will think in other time units. The answer to this initial open question generally provides the most appropriate time period to use.

In the questionnaire, a series of most commonly used time periods is provided so that during the interview as few calculations as possible are needed. Before the end of the interview, the enumerator will, however, convert the values reported for the different items to annual values. This is done by multiplication of the reported value with the multiplication factor given at the top of the schedule. Thus, if an item of information was given for a month (in column c), that value is multiplied by 12 to obtain the annual value in column (e).

Item	Code	*Multiplication Factor:* 365	365/7	12	4	
		Daily	**Weekly**	**Monthly**	**Quarterly**	**Annual**
		(a)	(b)	(c)	(d)	(e)

The totals for the year can then be calculated and the difference between income and expenditure, adjusted for stock changes, gives an indication of the unit's profit. As there are no external documents to authenticate the information given in the questionnaire (unlike the case of units maintaining accounts where the annual accounts provide a cross-check), it is very important that the implied profits are calculated. In general, profitability of small units is not very high as there often are few barriers to entry and it is also very unlikely that large losses are incurred as the financial capacity to absorb these is limited. Thus, the calculated profits provide a good indication of the quality of the response.

In cases where the resulting profit (or loss) seems to be implausible, this is first tested in a direct question to the respondent. The answer, if it clarifies the position, is noted down and

the response is complete. Otherwise, the easiest way of verifying the accuracy of the responses is by checking the largest items (on an annual basis they are directly comparable) amongst the incomes and expenditures and asking supplementary questions, such as the volatility of income or expenditure, or by using a different time frame to probe and resolve the problem.

Annex 1:   Industrial Survey – Large Questionnaire

INDUSTRIAL SURVEY - Concept   **Large**

1. IDENTIFICATION AND MAIN CHARACTERISTICS

a. IDENTIFICATION

Name of the Establishment _____ Establishment no: [____]

Address: _____

Telephone no: _____ Telefax No. _____

E-mail address: _____ Web-site Http:// _____

Contact person _____ Mobile No. _____

b. MAIN ACTIVITY

* The principle activity is identified as the activity which generates the highest income.   ISIC code

Specify the main activity in detail _____ [____]

Specify the secondary activities in detail   (a.) _____ [____]

                                             (b.) _____ [____]

                                             (c.) _____ [____]

c. ORGANIZATIONAL STATUS

101 Type of legal organization: *(Please tick)*	102 Type of ownership *(Please tick)*
[ ] 1. Individual proprietorship	[ ] 1. Privately owned
[ ] 2. Partnership	[ ] 2. Government owned
[ ] 3. Private Limited Company	[ ] 3. Government with private local collaboration
[ ] 4. Public Limited Company	[ ] 4. Foreign controlled
[ ] 5. Co-operative society	[ ] 5. Private with foreign collaboration
[ ] 6. Government department	[ ] 6. Government with foreign collaboration
[ ] 7. Other (specify)	[ ] 7. Other (specify)

2 Does the establishment have a parent company ?   Yes / No

If Yes, Name of parent company: _____

       Address of parent _____

       _____

3 Year of commencement of operations : [____]

4 Number of months in operation per year : [____] Months   SEASONAL : Yes / No

5 Does the unit maintain accounts ?   Yes / No

6 If Yes, accounting period over 12 months   From _____   to _____

Name of Statistical Service Field Staff:…………………………   Date of Interview: ……/……/………

Signature…………………………………..

2. EMPLOYMENT AND WAGES

Type of worker	Code	Number of workers			Remuneration paid to employees during the year				
		Average Number of workers during the year *(=b+c)*	*Of which:* Males	*Of which:* Females	Cash salaries and wages received (incl. bonuses and allowances)	Other benefits in cash or kind (e.g. food and housing benefits)	Social insurance premiums (incl. Health) withheld from the salary	Social premiums / Health insurance paid by the employer	**Total** remuneration paid during the period *(=d+e+f+g)*
Locals		(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)
Working Proprietors/partners	201								
Unpaid Family and other workers	202								
Apprentices	203								
Working Directors	204								
Permanent employees	205								
Casual/seasonal employees	206								
Leased workers **	207								
Total local staff (201 to 207)	209								
Foreigners									
Working Proprietors/partners	211								
Unpaid Family and other workers	212								
Apprentices	213								
Working Directors	214								
Permanent employees	215								
Casual/seasonal employees	216								
Leased workers **	217								
Total Foreign staff (211 to 217)	219								
Total all staff (209 and 219)	**299**								

Note: item 299(h) should equal item 435 in section 4

** Leased workers consist of (temporary) staff provided by outside employment agencies, etc.

** Wages and salaries cannot be provided for this category of staff. Instead, payments to the agencies providing the staff may be taken as approximation.

2. EMPLOYMENT AND WAGES – contd

Type of worker	Code	Number of workers			Remuneration paid to employees during the year				
		Average Number of workers during the year *(=b+c)*	*Of which:* Males	*Of which:* Females	Cash salaries and wages received (incl. bonuses and allowances)	Other benefits in cash or kind (e.g. food and housing benefits)	Social insurance premiums (incl. Health) withheld from the salary	Social premiums / Health insurance paid by the employer	**Total** remuneration paid during the period *(=d+e+f+g)*
Further detail of employment by:		(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)
Local Employees									
Operatives (production workers)	231								
Other workers	232								
All local employees	**239**								

Item 239 should equal items 205 plus 206 above

Type of worker	Code	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)
Foreign Employees									
Operatives (production workers)	241								
Other workers	242								
All foreign employees	**249**								

Item 249 should equal items 215 plus 216 above

Type of worker	Code	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)
Outworkers *(included under various categories of workers)*									
Local	251								
Foreign	252								
All outworkers	**259**								

Full-time and part-time workers

Local Workers *(all, including working proprietors, unpaid workers and employees)*

Type of worker	Code	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)
Full-time	261								
Part-time	262								
All local workers	**269**								
Foreign Workers *(all)*									
Full-time	271								
Part-time	272								
All foreign employees	**279**								

2.h. HOURS WORKED

Type of worker	Code	Number of hours worked during the year (=b+c)	By gender		By type		Full-time employee equivalents (=a/@)
			Males	Females	Operatives (production workers)	Other workers	
		(a)	(b)	(c)	(d)	(e)	(e)
Locals							
Working Proprietors/partners	201h						
Unpaid Family and other workers	202h						
Apprentices	203h						
Working Directors	204h						
Permanent employees	205h						
Casual/seasonal employees	206h						
Leased workers **	207h						
Total local staff (201h to 207h)	**209h**						
Foreigners							
Working Proprietors/partners	211h						
Unpaid Family and other workers	212h						
Apprentices	213h						
Working Directors	214h						
Permanent employees	215h						
Casual/seasonal employees	216h						
Leased workers **	217h						
Total Foreign staff (211h to 217h)	**219h**						
Total all staff (209h and 219h)	**299h**						

Number of hours in a regular full-time job per ☐ Week ☐ Month ☐ Year @

2.m EMPLOYMENT AND WAGES Male workers

Type of worker	Code	Number of workers			Remuneration paid to employees during the year				
		Average Number of workers during the year *(=b+c)*	*Of which:* Males	*Of which:* Females	Cash salaries and wages received (incl. bonuses and allowances)	Other benefits in cash or kind (e.g . food and housing benefits)	Social insurance premiums (incl. Health) withheld from the salary	Social premiums / Health insurance paid by the employer	**Total** remuneration paid during the period *(=d+e+f+g)*
Locals		(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)
Working Proprietors/partners	201m								
Unpaid Family and other workers	202m								
Apprentices	203m								
Working Directors	204m								
Permanent employees	205m								
Casual/seasonal employees	206m								
Total local staff (201m to 206m)	**209m**								
Foreigners									
Working Proprietors/partners	211m								
Unpaid Family and other workers	212m								
Apprentices	213m								
Working Directors	214m								
Permanent employees	215m								
Casual/seasonal employees	216m								
Total Foreign staff (211m to 213m)	219m								
Total all staff (209m and 219m)	**299m**								

Further detail of employees by:

Local Employees

Type of worker	Code								
Operatives (production workers)	231m								
Other workers	232m								
All local employees	**239m**								

Item239m should equal items 205m plus 206m above

Foreigner Employees

Type of worker	Code								
Operatives (production workers)	241m								
Other workers	242m								
All foreign employees	**249m**								

Item249m should equal items 215m plus 216m above

2.f EMPLOYMENT AND WAGES — Female workers

Type of worker	Code	Number of workers			Remuneration paid to employees during the year				
		Average Number of workers during the year (=b+c)	Of which: Males	Of which: Females	Cash salaries and wages received (incl. bonuses and allowances)	Other benefits in cash or kind (e.g. food and housing benefits)	Social insurance premiums (incl. Health) withheld from the salary	Social premiums / Health insurance paid by the employer	Total remuneration paid during the period (=d+e+f+g)
		(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)
Locals									
Working Proprietors/partners	201f								
Unpaid Family and other workers	202f								
Apprentices	203f								
Working Directors	204f								
Permanent employees	205f								
Casual/seasonal employees	206f								
Total local staff (201f to 206f)	**209f**								
Foreigners									
Working Proprietors/partners	211f								
Unpaid Family and other workers	212f								
Apprentices	213f								
Working Directors	214f								
Permanent employees	215f								
Casual/seasonal employees	216f								
Total Foreign staff (211f to 213f)	219f								
Total all staff (209f and 219f)	**299f**								

Further detail of employees by:

Local Employees

Type of worker	Code	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)
Operatives (production workers)	231f								
Other workers	232f								
All local employees	**239f**								

Item239f should equal items 205f plus 206f above

Foreigner Employees

Type of worker	Code	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)
Operatives (production workers)	241f								
Other workers	242f								
All foreign employees	**249f**								

Item249f should equal items 215f plus 216f above

65

3. INCOME

Item	Code	Value
Value of sales from own production *(finished products and by-products)* **	301	
** item 301 should equal item 1299(d) in section 12		
Value of sales of semi-finished goods	302	
Receipt from sale of used goods by the establishment	303	
(items such as sacks, drums etc.)		
Receipt for sale of waste and scrap products	304	
Receipts for contract work done *(on goods not owned by the company)*	305	
Receipts for repair and maintenance work done	306	
Value of own produced new construction, improvement or renovation made on fixed assets for own use	307	
** Item 307 should equal item 799(c) in Section 7		
Value of sales of goods bought and resold in the same condition as purchased	308	
** Item 308 should equal item 1801(a) in Section 18		
Value of sales of water	309	
** Item 309 should equal item 1004(b) in Section 10		
Value of sales of electricity	310	
** Item 310 should equal item 903(b) in Section 9		
Value of sales of wastewater treatment services	311	
Receipts for transport services rendered *(including charges for delivery of goods sold, when billed separately)*	312	
Income from commissions, etc.	313	
Income from rental of buildings	314	
Income from rental or lease of machinery and equipment	315	
Income from the sale or licensing of R & D products	316	
** Item 316 should equal item 1669 in Section 16		
Income from rental of land	317	
Interest received	318	
Dividends received	319	
Other earnings on investments *(Please specify)*		
	320	
	321	
Subsidies received	322	
Export rebates received	323	
Gains on sales of fixed assets	324	
Receipts from insurance claims	325	
Exchange gains	326	
Other Income *(Please specify)*		
	327	
	328	
	329	
Total Income (sum of items number 301 to 329)	**399**	

(Please transfer item 399 to item 602 in Section 6)

4. EXPENDITURE

Item	Code	Value
Payments made for raw materials, chemicals, packing materials and spare parts	401	
**Item 401 is equal to item 1199(d) in section 11		
Payments made to other firms and individuals for work performed on sub-contract	402	
Payments made for repair and maintenance work	403	
Payment made for fuel purchased (not for household use)	404	
** Item 404 is equal to item 819(d) in Section 8		
Payment made for electricity purchased (not for household use)	405	
** Item 405 is equal to item 901(b) in Section 9		
Payment made for water purchased (not for household use)	406	
** Item 406 is equal to item 1001(b) in Section 9		
Purchase of goods for resale (traded goods)	407	
** Item 407 is equal to item 1811(a) in Section 18		
Commissions Paid	408	
Payment for rental of warehouses, industrial space, etc.	409	
Payments for rental of office, shops & other premises	410	
Payments for Rental or Lease of Machinery and Equipment	411	
Payments for Installation Charges	412	
Transportation charges paid	413	
Payments made for telephone	414	
Payments made for internet access	415	
Payments made for postage, couriers, etc.	416	
Payments made for printing and stationery	417	
Purchases of Books and Journals	418	
Payments of travelling costs for business purposes	419	
Payment of Insurance Premiums	420	
Bank Charges (excluding bank interest charges)	421	
Payments made for advertising and promotion (excluding printing)	422	
Payment of Management Fees	423	
Payments Made for Legal Expenses	424	
Payments for Auditing and Accountants' Fees	425	
Payments for Other Professional Services	426	
Payments made for Licence/registration fees	427	
Payment of royalties, copyrights, patents, etc.	428	
Payments for education and training	429	
Cost of Treatment and Disposal of Waste Products	430	
Depreciation charges for the year on fixed assets	431	
** item 431 should equal item 799(h) in section 7		
Interest Paid	432	
Payments for Rent of Land	433	
Payments of taxes on production	434	
(e.g. Sales Tax, Excise Tax, Hotel Tax, Mineral Duty, Timber Levy etc.)		
Payments made for wages and salaries	435	
** item 435 should equal item 299(h) in section 2		
Directors' Remuneration, Attendance fees and Allowances	436	
Provision for Bad or Doubtful Debts	437	
Loss from sales of fixed assets	438	
Loss from investments and stocks	439	
Exchange losses	440	
Cash donations, grants & gifts	441	
Extraordinary items *(please specify)* *(This should include all items in the annual accounts not specified above)*		
	442	
	443	
	444	

5. STOCKS AND INVENTORIES

Stocks of:	Code	(+) Closing Stock (a)	(-) Opening Stock (b)	(=) Increase in Stocks (c)
Raw Materials	501			
Chemicals and dyes	502			
Packing material	503			
Spare parts	504			
Fuels and Lubricants	505			
Work in progress	506			
Finished Goods	507			
Goods purchased for Resale	508			
Any other stocks and inventories	509			
Total	**599**			

Please transfer item 599(a) to item 605(b) in Section 6.

Please transfer item 599(b) to item 601(b) in Section 6.

Work in progress includes the value of output produced by your enterprise but not yet in the state in which it is normally sold.

6. PROFITS AND THEIR APPROPRIATION

Description	Code	Detail Values (a)	Item Values (b)	(Sub) Totals (c)
Value to stocks at end of accounting year (copy item 599(a) in Section 5)	601			
Total income (copy item 399 in Section 3)	602			
Sub-Total (601 + 602)	603			
Total expenditure (Copy item 499 in Section 4)	604			
Value of stocks at beginning of accounting year (Copy item 599(b) in Section 5)	605			
Sub-Total (604 + 605)	606			
Profit before taxation @ Sub-Total (603 – 606)	607			
Unappropriated profits brought forward from previous year @	608			
Sub-Total (607 + 608)	609			
Profit retained in business @	610			
Profit distributed as dividends, etc. @	611			
of which: Profits distributed to foreign shareholders	*611.1*			
Profits distributed to local shareholders	*611.2*			
Sub-Total (610 + 611)	612			
Unappropriated profits carried forward to following year @	613			

(equals 609 – 612)

@ Items 607, 608, 610, 611, 613 should equal the same items reported in the annual accounts of the establishment.

7. FIXED ASSETS

Item	Code	(+) Book Value of assets at start of the year (a)	(+) Purchases of fixed assets (b)	(+) Own account capital formation (c)	(+) Revaluations of fixed assets (d)	(-) Sales of fixed assets (e)	(-) Fixed assets lost or destroyed (f)	(=) Gross additions to fixed assets (g) = b+ c+d e-f	(-) Depreciation charges for the year (h)	(=) Book value of assets at end of the year (i)= a+g–h	Depre-ciation rates used (j)
Land	701										
Land Improvement	702										
Residential Buildings	703										
Other Buildings	704										
Other structures, e.g. roads, bridges, mine shafts, etc.	705										
Transport equipment @	706										
Computers and related equipment @	707										
Other Plant and Machinery @	708										
Capital work in progress	709										
Livestock for breeding, dairy, draught, etc	710										
Vineyards, orchards and other plantations of trees yielding repeat products	711										
Intellectual property products	712										
Mineral exploration	713										
Computer software and databases	714										
Entertainment, literary or artistic originals	715										
Other intangible fixed assets (Specify)											
	716										
	717										
Total	**799**										

@ Include machinery under financial lease, I.e. where lease payments over the contract life will equal its value

(1) Net additions to the value of assets recorded in the organisation's balance sheet due to increases in the price of assets owned

** Item 799(h) should be equal to item 431 in Section 4

** Item 799(c) should be equal to item 307 in Section 3

8. FUEL PURCHASED AND CONSUMED

Item Description .	Code	Unit of measurement	Quantity (a)	Factor (b)	Tera Joules (c) = (a) * (b)	Value (d)
Solid fuels						
Fuelwood	801	Metric ton				
Charcoal	802	Metric ton				
Coal	803	Metric ton				
Coke	804	Metric ton				
Liquid fuels						
High speed diesel oil	805	1000 litre				
Light diesel oil	806	1000 litre				
Gasoline/petrol	807	1000 litre				
Kerosene	808	1000 litre				
Furnace oil	809	1000 litre				
Engine oil	810	1000 litre				
Liquefied gas (LPG / CNG)	811	Metric ton				
Gaseous fuels						
Natural gas / factory (coking) gas	812	1000 Cu Mtr				
Other fuels & lubricants *(please specify)*						
	813					
	814					
	815					
Total fuels purchased (801 to 815)	**819**	Tera Joule				
Minus changes in stocks of:						
Fuels and Lubricants	821	*(same as item 505(c) from section 5)*				
Plus fuels recovered from waste						
Methane recovered from waste products *(this should equal item 1052 in section 10)*	822	1000 Cu Mtr				
Heat recovered from incineration of waste *(this should equal item 1054 in section 10)*	823	Tera Joule				
Total fuels consumed (819 to 823)	**829**					

Note: Fuel used in manufacturing processes and generation of electricity should also be reported here.

Total purchased fuels used (819 and 821)	**899**					

Note: item 821(c), if not available, can be estimated using the ratio of items 819(c) / 819(d)

8.a. Of which: From renewable sources

Item Description	Code	Unit	Quantity	Factor	Tera Joules	Value
Solid fuels						
Fuelwood	851	Metric ton				
Charcoal	852	Metric ton				
Liquid fuels						
Bio diesel	853	1000 liters				
Ethanol	854	1000 liters				
Other fuels & lubricants *(please specify)*						
	855					
	856					
Methane recovered from waste products *(this should equal item 1052 in section 10)*	851	1000 Cu Mtr				
Heat recovered from incineration of waste *(this should equal item 1054 in section 10)*	852	Tera Joule				
Total from renewable sources	**859**	Tera Joule				
Percentage from renewable sources		*(item 859(c) / 899(c))*		%		

9. ELECTRICITY GENERATED, PURCHASED AND USED

Item Description	Code	Unit of measurement	Quantity (a)	Value (b)
Electricity purchased (or obtained from other establishments of the same company)	901	1000 KWH		
Electricity generated	902	1000 KWH		
Electricity sold	903	1000 KWH		
Total Electricity used (901 + 902- 903)	**999**	**1000 KWH**		

Item 901(b) is the same as item 404 in Section 4.

Item 903(b) is the same as item 310 in Section 3.

9.a. Of which: Generated from renewable sources

Item Description	Code	Unit of measurement	Quantity (a)	Value (b)
Hydro	951	1000 KWH		
Solar	952	1000 KWH		
Wind	953	1000 KWH		
Biomass/biofuels @	954	1000 KWH		
Other renewables (specify)				
	955	1000 KWH		
	956	1000 KWH		
Total electricity generated from renewable sources	**959**	**1000 KWH**		

@ "Biofuels are fuels of biological and renewable origin, such as fuelwood, charcoal, lifestock manure, biogas, biohydrogen, bioalcohol, microbial biomass, agricultural waste and byproducts, energy crops, and others." from FAO http://www.fao.org/sd/EGdirect/EGre0055.htm

Percentage from renewable sources   (item9859(c) / 999(c))   %

10. WATER PRODUCED, PURCHASED AND USED

Item Description	Code	Unit of measurement	Quantity (a)	Value (b)
Water purchased	1001	1000 litre		
Water produced / pumped up / collected / drawn in	1002	1000 litre		
Water reused / recycled	1003	1000 litre		
Water sold	1004	1000 litre		
Total Water used (1001 to 1003 – 1004)	**1099**	**1000 litre**		

Item 1001(b) is the same as item 406 in Section 4.
Item 1004(b) is the same as item 309 in Section 3.

10.a. WASTE TREATMENT

Industrial wastewater treated (to local environmental norms)	1051	1000 litre		
Recovery of gas (methane)/fermentation	1052	1000 Cu Mtr		
Composting	1053	1000 Tonnes		
Incineration of solid waste/heat recovery	1054	1000 Tonnes		
Treated water reused/recycled	1055	1000 litre		

11. GOODS & MATERIALS CONSUMED (not bought for resale in the same condition)

Description of raw materials used	Code	Commodity Code (for official use only) (a)	Unit (b)	Total Use		Items of which imported directly	
				Quantity (c)	Value (d)	Quantity (e)	Value (f)
1	1101						
2	1102						
3	1103						
4	1104						
5	1105						
6	1106						
7	1107						
8	1108						
9	1109						
10	1110						
11	1111						
12	1112						
13	1113						
14. All other materials used	1114						
Total value of materials used	**1149**						

Minus changes in stocks of:

Raw Materials	1151	*(item 501(c) from section 5)*
Chemicals and dyes	1152	*(item 502(c) from section 5)*
Packing material	1153	*(item 503(c) from section 5)*
Spare parts	1154	*(item 504(c) from section 5)*
Any other stocks and inventories	1155	*(item 509(c) from section 5)*
Total decrease in stocks	1159	*(sum of items 1151 to 1155)* **-/-**
Total value of materials purchased	**1199**	*(item 1149 minus item 1159)* This should equal item 401 in section 4

Note: These inputs and materials should be valued at the cost actually incurred in the purchase of them.

12. GOODS PRODUCED (not sold in the same condition as purchased)

Description of goods produced	Code	Commodity Code (for official use only)	Unit	Total Production		Items of which Exported directly	
				Quantity	Value	Quantity	Value
		(a)	(b)	(c)	(d)	(e)	(f)
1	1201						
2	1202						
3	1203						
4	1204						
5	1205						
6	1206						
7	1207						
8	1208						
9	1209						
10	1210						
11	1211						
12	1212						
13	1213						
14. All other goods produced	1214						
Total value of production	**1249**						

Plus changes in stocks of:

Work in progress	1251	*(copy item 506(c) from section 5)*	
Finished Goods	1252	*(copy item 507(c) from section 5)*	**+**
Total value of sales	**1289**		
Minus:			
Value of sales of semi-finished goods	1291	*(copy item 302 from section 3)*	**-/-**
Total value of sales of finished goods	**1299**		This should equal item 301 in section 3

Note: All products should be valued at the expected sales prices.

13. PRODUCTION CAPACITY AND ITS UTILISATION

Description of product / lines (Please use -to the extent possible- the same order as in Section 12)	Code	Commodity code (for official use only)	Unit of measurement	Total Production capacity	Percentage capacity used during the year	Reason for low utilisation
		(a)	(b)	(c)	(d)	(e)
1	1301					
2	1302					
3	1303					
4	1304					
5	1305					
6	1306					
7	1307					
8	1308					
9	1309					
10	1310					
11	1311					
12	1312					
13	1313					

14. SALES AND PROFITABILITY

Item	Code	Much better	Some-what better	About the same	Some-what worse	Much worse
		(a)	(b)	(c)	(d)	(e)
Sales						
Compared with the previous year, how are your sales of this year	1401					
Compared with four years ago, how are this year's sales	1402					
Profits						
Compared with the previous year, how are your profits for this year	1403					
Compared with four years ago, how are this year's profits	1404					

Remarks:

15. GENERAL BUSINESS ENVIRONMENT

Problem Category	Code	Very serious Problem	Serious problem	Minor problem	No problem	Not Applicable	Compared with four years ago, has the situation	
							Improved	Deteriorated
		(a)	(b)	(c)	(d)	(e)	(f)	(g)
Availability of affordable finance / bank loans	1501							
Availability of tools and machinery	1502							
Availability of working space	1503							
Availability of skilled/technical labour	1504							
Availability of unskilled labour	1505							
Availability of electricity	1506							
Availability of water	1507							
Availability of input materials	1508							
Availability of transport facilities	1509							
Government/regulatory environment	1510							
Marketing problems	1511							
Other problems *(please specify)*								
	1512							
	1513							
	1514							
	1515							
	1516							

16. RESEARCH AND DEVELOPMENT

16.a Number of R & D workers		Number of workers			Remuneration paid to employees during the year				
Number of workers engaged in research and development	Code	Average Number of workers during the year *(=b+c)*	*Of which:* Males	*Of which:* Females	Cash wages and salaries received (incl. bonuses and allowances)	Other benefits in cash or kind (e.g. food and housing benefits)	Social insurance premiums (incl. Health) withheld from the salary	Social premiums / Health insurance paid by the employer	**Total** remuneration paid during the period *(=d+e+f+g)*
		(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)
Intellectual property products	1601								
Mineral exploration and evaluation	1602								
Computer software and databases	1603								
Research and development	1604								
Entertainment, literary and artistic originals	1605								
All workers in R & D	**1649**								

16.b Value of research work capitalised during the year

Item	Code	Value	
Intellectual property products	1651		This item should be transferred to item 712(c) in Section 7
Mineral exploration and evaluation	1652		This item should be transferred to item 713(c) in Section 7
Computer software and databases	1653		This item should be transferred to item 714(c) in Section 7
Research and development	1654		This item should be transferred to item 712(c) in Section 7
Entertainment, literary and artistic originals	1655		This item should be transferred to item 715(c) in Section 7
Total value of work capitalised	**1659**		

16.c Value of research work sold or licensed during the year

Item	Code	Value	
Intellectual property products	1661		
Mineral exploration and evaluation	1662		
Computer software and databases	1663		
Research and development	1664		
Entertainment, literary and artistic originals	1665		
Total value of work sold or licensed	**1669**		This item should be transferred to item 316 in Section 3

17. ORDERS

Description of goods produced **(Please use the same order as in Section 12)**	Code	Commodity code (for official use only)	Unit of measure-ment	Unfilled orders at the beginning of the year	New orders received during the year	Orders filled during the year	Orders cancelled during the year	Unfilled orders at the end of the year =(c+d-e-f)
		(a)	(b)	(c)	(d)	(e)	(f)	(g)
1	1701							
2	1702							
3	1703							
4	1704							
5	1705							
6	1706							
7	1707							
8	1708							
9	1709							
10	1710							
11	1711							
12	1712							
13	1713							
14. All other goods	1714							
Total value of orders	**1799**							

18. TRADED GOODS (goods sold in the same condition as purchased)

	Code	Detail (a)	Value (b)
Sales of goods bought and resold in the same condition (traded goods)	1801		
** Item 1801(a) should equal item 308 in Section 3			
Goods withdrawn for own use (if any)	1802		
Goods given to employees (if any)	1803		
Closing stock at the <u>end</u> of the year	1804		
** Item 1804(a) should equal item 508(b) in Section 5			
Total use of goods purchased for resale *(1801 to 1804)*	1809		
Purchase of goods for resale in the same condition (traded goods)	1811		
** Item 1811(a) should equal item 407 in Section 4			
Opening stock at the <u>beginning</u> of the year	1812		
** Item 1812(a) should equal item 508(a) in Section 5			
Total supply of goods purchased for resale *(1811 and 1812)*	1819		
Gross trade margin *(1809 – 1819)*	1899		
AVERAGE Percentage margin added to purchase price *(item 1899 / item 1819)*	1851	%	

Annex 2: Industrial Survey – Small Questionnaire

INDUSTRIAL SURVEY - Concept   **Small**

1. IDENTIFICATION AND MAIN CHARACTERISTICS

a. IDENTIFICATION

Name of the Establishment _____ Establishment no: ☐

Address: _____

Telephone no: _____ Telefax No. _____

E-mail address: _____ Web-site Http:// _____

Contact person _____ Mobile No. _____

b. MAIN ACTIVITY

*The principle activity is identified as the activity which generates the highest income.   ISIC code

Specify the main activity in detail _____ ▊

Specify the secondary activities in detail   (a.) _____ ▊

(b.) _____ ▊

(c.) _____ ▊

c. ORGANIZATIONAL STATUS

101 Type of legal organization: *(Please tick)*   102   Type of ownership *(Please tick)*

☐	1. Individual proprietorship	☐	1. Privately owned
☐	2. Partnership	☐	2. Government owned
☐	3. Private Limited Company	☐	3. Government with private local collaboration
☐	4. Public Limited Company	☐	4. Foreign controlled
☐	5. Co-operative society	☐	5. Private with foreign collaboration
☐	6. Government department	☐	6. Government with foreign collaboration
☐	7. Other (specify)	☐	7. Other (specify)

2   Does the establishment have a parent company ?   Yes / No

If Yes, Name of parent company: _____

Address of parent _____

_____

3   Year of commencement of operations :   ☐

4   Number of months in operation per year : ☐ Months   SEASONAL :   Yes / No

5   Does the unit maintain accounts ?   Yes / No

6   If Yes, accounting period over 12 months   From _____ to _____

Name of Statistical Service Field Staff:…………………………   Date of Interview: ……/……/……….

Signature……………………………………..

82

2. EMPLOYMENT AND WAGES

Employment data relate to: **a. Past year**   **b. One month, namely _____, 20...**

Type of worker	Code	Average number of workers during the period *(=b+c)*	*Of which:* Males	*Of which:* Females	**Total** wages and salaries paid during the month	**Total** wages and salaries paid during the year
Locals		(a)	(b)	(c)	(d)	(e)
Working Proprietors/partners	201					
Unpaid Family and other workers	202					
Apprentices	203					
Working Directors	204					
Permanent employees	205					
Casual/seasonal employees	206					
Total local staff (201 to 206)	209					
Foreigners						
Working Directors	211					
Permanent employees	215					
Casual/seasonal employees	216					
Total Foreign staff (211 to 213)	219					
Total all staff (209 and 219)	**299**					

Further detail of employees by: Note: item 299(e) should equal item 435(e) in section 4

Local Employees

Type	Code	(a)	(b)	(c)	(d)	(e)
Operatives	231					
Other workers	232					
All local employees	**239**					

Item 239 should equal items 205 plus 206 above

Foreigner Employees

Type	Code	(a)	(b)	(c)	(d)	(e)
Operatives	241					
Other workers	242					
All foreign employees	**249**					

Item 249 should equal items 212 plus 213 above

Note: If monthly information is reported, annual data are obtained by multiplying the monthly data by 12.

3. INCOME

Item	Code	Daily (a)	Weekly (b)	Monthly (c)	Quarterly (d)	Annual (e)
Multiplication Factor:		*365*	*365/7*	*12*	*4*	
Value of sales from own production (finished products and by-products) **	301					
** item 301(e) should equal item 1299(h) in section 12						
Value of sales of semi-finished goods	302					
Receipt from sale of used goods by the establishment	303					
(items such as sacks, drums etc.)						
Receipt for sale of waste and scrap products	304					
Receipts for contract work done (on goods not owned by he company)	305					
Receipts for repair and maintenance work done	306					
Value of own produced new construction, improvement or renovation made on fixed assets for own use	307					
** Item 307(e) should equal item 799(c) in Section 7						
Value of sales of goods bought and resold in the same condition as purchased	308					
** Item 308(e) should equal item 1801(a) in Section 18						
Receipts for transport services rendered *(including charges for delivery of goods sold, when billed separately)*	309					
Income from commissions, etc.	310					
Income from rental of buildings	311					
Income from rental or lease of machinery and equipment	312					
Income from the sales of other services not specified above	313					
Income from rental of land	314					
Other Income *(Please specify)*						
	351					
	352					
	353					
Total Income (sum of items number 301 to 353)	**399**					

(Please transfer item 399(e) to item 602 in Section 6)

Note: For each item, the most convenient reporting period is used and given in any of the columns (a) to (e)..

If the value reported is not for the year, the annual value is calculated by using the multiplication factors given at the top of each of the columns (a) to (d).

4. EXPENDITURE

Item	Code	Daily (a)	Weekly (b)	Monthly (c)	Quarterly (d)	Annual (e)
Multiplication Factor:		*365*	*365/7*	*12*	*4*	
Payments made for raw materials, chemicals, packing materials and spare parts **Item 401(e) is equal to item 1199(h) in section 11	401					
Payments made for repair and maintenance work	403					
Payment made for fuel purchased (not for household use) ** Item 404(e) is equal to item 819(h) in Section 8	404					
Payment made for electricity purchased (not for household use) ** Item 405(e) is equal to item 901(h) in Section 9	405					
Payment made for water purchased (not for household use) ** Item 406(e) is equal to item 1001(h) in Section 9	406					
Purchase of goods for resale (traded goods) ** Item 407(e) is equal to item 1811(a) in Section 18	407					
Payments for rental of buildings	451					
Payments for rental or lease of machinery and equipment	411					
Transportation charges paid	413					
Payments made for telephone, internet, postage, courier	452					
Payments made for printing, stationery	417					
Payments of travelling costs for business purposes	419					
Payment of Insurance Premiums	420					
Payments for Auditing and Accountants' Fees and other professional services	453					
Payments made for Licence/registration fees	427					
Depreciation charges for the year on fixed assets ** item 431(e) should equal item 799(h) in section 7	431					
Payments for wages and salaries *(435(e) should equal item 299(g) in section 2)*	435					
Other items *(These should include all current expenditures not specified above) (please specify)*						
	454					
	455					
	456					
Total Expenditures (sum of items number 401 to 456)	**499**					

(Please transfer item 499(e) to item 604 in Section 6)

Note: For each item, the most convenient reporting period is used and given in any of the columns (a) to (e)..

If the value reported is not for the year, the annual value is calculated by using the multiplication factors given at the top of each of the columns (a) to (d).

5. STOCKS AND INVENTORIES
Note: if stock levels are small or changes in stocks are small, this section may be ignored

Stocks of:	Code	Stock (a)	Stock (b)	Stocks (c)
Raw materials, chemicals, spare parts, etc.	551			
Fuels and Lubricants	505			
Work in progress @	506			
Finished Goods	507			
Goods purchased for Resale	508			
Total	**599**			

Please transfer item 599(a) to item 605(b) in Section 6.

Please transfer item 599(b) to item 601(b) in Section 6.

@ Work in progress includes the value of output produced by your enterprise but not yet in the state in which it is normally sold.

6. PROFITS AND THEIR APPROPRIATION

Description	Code	Detail Values (a)	Item Values (b)	(Sub) Totals (c)
Value to stocks at end of accounting year (copy item 599(a) in Section 5)	601			
Total income (copy item 399(e) in Section 3)	602			
Sub-Total (601 + 602)	603			
Total expenditure (Copy item 499(e) in Section 4)	604			
Value of stocks at beginning of accounting year (Copy item 599(b) in Section 5)	605			
Sub-Total (604 + 605)	606			
Profits / losses **(603 − 606)**	**699**			

If a **loss** is reported for the period, please give an explanation:

_____

_____

14. SALES AND PROFITABILITY

Item	Code	Much better (a)	Some-what better (b)	About the same (c)	Some-what worse (d)	Much worse (e)
Sales						
Compared with the previous year, how are your sales this year	1401					
Compared with four years ago, how are this year's sales	1402					
Profits						
Compared with the previous year, how are your profits for this year	1403					
Compared with four years ago, how are this year's profits	1404					

Remarks:

7. FIXED ASSETS

Item	Code	Book Value of assets at beginning of the year (a)	(+) Purchases of fixed assets (b)	(+) Own account capital formation (c)	(-) Sale of fixed Assets (e)	(=) Gross additions to fixed assets (g)	(-) Depreciation charges for the year (h)	(=) Book value of assets at end of the year (i) = (g) – (h)	Depre-ciation rates used (j)
Land	701								
Land Improvement	702								
Residential Buildings	703								
Other Buildings	704								
Other structures, e.g. Roads, Bridges, Mine shafts, etc.	705								
Transport equipment @	706								
Computers and related equipment @	707								
Other Plant and Machinery @	708								
Livestock for breeding, dairy, draught, etc.	709								
Vineyards, orchards and other plantations of trees yielding repeat products	710								
Mineral exploration	711								
Computer software and databases	712								
Entertainment, literary or artistic originals	713								
Other intangible fixed assets (Specify)									
	714								
	715								
Total	**799**								

 Include machinery under financial lease, i.e. where lease payments over the contract life will equal its value

(1) Net additions to the value of assets recorded in the organisation's balance sheet due to increases in the price of assets owned

   ** Item 799(h) should be equal to item 431 in Section 4

   ** Item 799(c) should be equal to item 307 in Section 3

8. FUEL PURCHASED AND CONSUMED

Item	Code	Unit (a)	Quantity (b)	Annual Quantity (c)	Total value of purchases of fuels: Daily 365 (d)	Weekly 365/7 (e)	Monthly 12 (f)	Quarterly 4 (g)	Annual Value (h)
				Multiplication Factors:	*365*	*365/7*	*12*	*4*	
Solid fuels									
Fuelwood	801	Metric ton							
Charcoal	802	Metric ton							
Coal	803	Metric ton							
Liquid fuels									
High speed diesel oil	804	1000 litre							
Light diesel oil	805	1000 litre							
Gasoline/petrol	806	1000 litre							
Kerosene	807	1000 litre							
Furnace oil	808	1000 litre							
Engine oil	809	1000 litre							
Liquefied gas (LPG / CNG)	810	Metric ton							
Gaseous fuels									
Natural gas / factory (coking) gas	811	1000 Cu Mtr							
Other fuels & lubricants *(please specify)*									
	812								
	813								
	814								
Total fuels purchased (801 to 814)	**819**								

Item 819(h) is equal to item 404(e) in Section 8

Note: For each fuel purchased, the most convenient reporting period is used.

The value of purchases is reported in columns (e) to (i) while the quantity purchased for the same period is reported in column (b).

If the value reported is not for the year, the annual value of purchases is calculated by using the multiplication factors given at the top of each of the columns (e) to (h).

The same multiplication factor is also used on the quantity reported in column (b) to obtain the annual quantity of purchases in column .

9. ELECTRICITY GENERATED PURCHASED AND USED

Item Description .	Code	Unit	Quantity (b)	Annual Quantity (c)	Daily (d)	Weekly (e)	Monthly (f)	Quarterly (g)	Annual Value (h)
			Multiplication Factors:		*365*	*365/7*	*12*	*4*	Total value:
Electricity purchased (or obtained from other establishments of the same company)	901	1000 KWH							
Electricity generated	902	1000 KWH							
Total Electricity used (901 + 902)	**999**	**1000 KWH**							

Item 901(h) is the same as item 404 in Section 4.

Note: For each item, the most convenient reporting period is used.

The values are reported in the relevant columns (d) to (h) while the quantity purchased for the same period is reported in column (b).

If the value reported is not for the year, the annual value for the item is calculated by using the multiplication factors given at the top of each of the columns (d) to (g).

The same multiplication factor is also used on the quantity reported in column (b) to obtain the annual quantity of purchases in column (c).

10. WATER PRODUCED, PURCHASED AND USED

Item Description .	Code	Unit	Quantity (b)	Annual Quantity (c)	Daily (d)	Weekly (e)	Monthly (f)	Quarterly (g)	Annual Value (h)
			Multiplication Factors:		*365*	*365/7*	*12*	*4*	Total value:
Water purchased	1001	1000 litre							
Water produced / pumped up	1002	1000 litre							
Water reused / recycled	1003	1000 litre							
Total Water used (1001 to 1003)	**1099**	**1000 litre**							

Item 1001(h) is the same as item 406 in Section 4.

Note: For each item, the most convenient reporting period is used.

The values are reported in the relevant columns (d) to (h) while the quantity purchased for the same period is reported in column (b).

If the value reported is not for the year, the annual value for the item is calculated by using the multiplication factors given at the top of each of the columns (d) to (g).

The same multiplication factor is also used on the quantity reported in column (b) to obtain the annual quantity of purchases in column (c).

11. PURCHASES OF GOODS & MATERIALS CONSUMED (not bought for resale)

Description of goods and raw materials purchased	Code	Commodity Code (for official use only) (a)	Quantity (b)	Annual Quantity (c)	Unit (d)	Total Value of Purchases of raw materials, etc. Multiplication Factors: 365 Daily (e)	365/7 Weekly (f)	12 Monthly (g)	4 Quarterly (h)	Annual Value (i)
1	1101									
2	1102									
3	1103									
4	1104									
5	1105									
6	1106									
7	1107									
8. All other materials purchased	1108									
Total value of all materials	**1199**									

Item 1199(h) should equal item 401 in section 4

12. SALES OF GOODS PRODUCED (not sold in the same condition as purchased)

Description of goods sold	Code	Commodity Code for official use only) (a)	Quantity (b)	Annual Quantity (c)	Unit (d)	Total value of sales of goods produced Multiplication Factors: 365 Daily (e)	365/7 Weekly (f)	12 Monthly (g)	4 Quarterly (h)	Annual (i)
1	1201									
2	1202									
3	1203									
4	1204									
5	1205									
6	1206									
7	1207									
8. All other goods sold	1208									
Total value of sales	**1299**									

Item 1299 (i) should equal item 301(e) in section 3

Note: For each product purchased or sold, the most convenient reporting period is used.

The value of purchases/sales is reported in columns (e) to (i) while the quantity of purchases/sales for the same period is reported in column (b).

If the value reported is not for the year, the annual purchase/sales value is calculated by using the multiplication factors given at the top of each of the columns (e) to (h).

The same multiplication factor is also used on the quantity reported in column (b) to obtain the annual quantity of purchase/sales in column (c).

90

15. GENERAL BUSINESS ENVIRONMENT

Problem Category	Code	Very serious problem	Serious problem	Minor problem	No problem	Not applicable	Compared with four years ago, has the situation	
							Improved	Deteriorated
		(a)	(b)	(c)	(d)	(e)	(f)	(g)
Availability of affordable finance / bank loans	1501							
Availability of tools and machinery	1502							
Availability of working space	1503							
Availability of skilled/technical labour	1504							
Availability of unskilled labour	1505							
Availability of electricity	1506							
Availability of water	1507							
Availability of input materials	1508							
Availability of transport facilities	1509							
Government/regulatory environment	1510							
Marketing problems	1511							
Other problems *(please specify)*								
	1512							
	1513							
	1514							
	1515							
	1516							

18. TRADED GOODS (goods sold in the same condition as purchased)

	Code	Detail (a)	Value (b)
Sales of goods bought and resold in the same condition (traded goods) ** Item 1801(a) should equal item 308 in Section 3	1801		
Goods withdrawn for own use (if any)	1802		
Goods given to employees (if any)	1803		
Closing stock at the <u>end</u> of the year ** Item 1804(a) should equal item 508(b) in Section 5	1804		
Total use of goods purchased for resale *(1801 to 1804)*	1809		
Purchase of goods for resale in the same condition (traded goods) ** Item 1811(a) should equal item 407 in Section 4	1811		
Opening stock at the <u>beginning</u> of the year ** Item 1812(a) should equal item 508(a) in Section 5	1812		
Total supply of goods purchased for resale *(1811 and 1812)*	1819		
Gross trade margin *(1809 – 1819)*	1899		
AVERAGE Percentage margin added to purchase price	1851	%	

Chapter 3: Practical Sampling Guidelines for Economic Surveys in Developing Countries

1.0   Introduction

All countries require various types of information on each sector of the economy for monitoring economic indicators, planning and developing economic policies and programmes. Some of this information can be obtained from different administrative sources, including the government accounts and from commissions or boards responsible for special subsectors of the economy. However, given the enormous size of most economic activities in the different sectors of the economy, the majority of countries conduct an Economic Census every five or ten years, followed by Economic Surveys on an annual, quarterly, and sometimes monthly basis relying on   nationally-representative samples of establishments.

The main objective of the Economic Census is to update the structural information on all sectors of the economy and to serve as a sampling frame for the annual, quarterly and monthly Economic Surveys. Given the heavy cost of conducting an Economic Census and the dynamic nature of most economies, information on the value of capital investment, employment, annual production or output, inputs, value-added and other measures is updated through periodic Economic Surveys conducted using a sample of establishments.

The purpose of these guidelines is to describe different methodologies that can be considered for developing the sampling frames and designing the samples for Economic Surveys. Procedures for updating the sampling frames are also described. Whilst the publication is mainly designed for statisticians working with establishment surveys in developing countries, others may benefit from the discussion of the various conceptual issues.

In some countries, the Economic Surveys are referred to as Establishment Surveys, given that the unit of data collection and analysis is generally the establishment. Many countries conduct multi-sectoral Economic Surveys, while others specialise in specific sectors, such as industry. The sampling procedures described in these guidelines can also be used for establishment surveys covering specific economic sectors. The establishment-based surveys are differentiated from household surveys, which are an important source of socioeconomic information collected through a sample of households. Some household surveys such as a

Household Income and Expenditure Survey (HIES) or living conditions survey are also used to measure the consumption components of the national accounts. Given that some informal economic activities are household-based, some countries also rely on household surveys to cover such activities.

In view of the importance of agriculture to the economies of many developing countries, as well as the unique nature of this sector, with many small farming operations throughout the country operated at a household level, most countries conduct separate agricultural surveys. Many countries have separate agricultural surveys for crop forecasting for food security and planning for imports and exports of agricultural products, as well as postharvest surveys for more accurate estimates of agricultural and livestock production needed for national accounts. Such agricultural surveys sometimes measure the cost of production, inputs and use of improved technologies to monitor different aspects of the agricultural sector.

Most developing countries use a sample household survey approach to the collection of data on crop and livestock production, using a sampling frame of primary sampling units (PSUs) such as census enumeration areas (EAs), with a new listing of farm households in sample EAs and a second stage selection of farm households. This multi-stage sampling approach is sometimes complemented by a list frame of large commercial farms that are included in the surveys with certainty given their relatively larger contribution to agricultural production. Some more developed countries conduct agricultural surveys based on a directory of farms similar to a Business Register. Agricultural surveys will not be covered in these guidelines which focus on the industrial sector.

This publication is designed to be practical, with examples of applications in different countries, and has limited discussion of statistical theory. However, it is helpful to begin with an overview of the basic sampling concepts.

2.0   Overview of Sampling Concepts

In order to better understand the various sampling approaches discussed in this publication, a basic description of the different sampling concepts is provided. The sampling methodologies discussed are related to probability or scientific samples, where all the units (establishments) in the sampling frame have a positive and known probability of selection. This type of sample allows us to make unbiased estimates of the parameters (true values) of the indicators for the entire population, and to measure the level of precision (sampling errors and confidence intervals). Other types of non-probability sampling, such as purposive (judgmental) samples and quota samples, are used for other purposes, such as conducting a pre-test but such methods do not allow us to make inferences (estimates) for the

population. One exception is cut-off sampling, which is discussed briefly in this report since it is sometimes used as a cost-effective way of obtaining indicators based on the largest establishments that cover a large percentage of the production and employment in specific sectors.

Since these guidelines are designed to cover practical sampling applications for Economic Surveys, only the main sampling concepts will be described here.[1]

2.1   Sampling frame and sampling units

The sampling frame is a list or database containing all the sampling units in a population such as economic enterprises or establishments. The type of sampling unit depends on the stages of selection in the sample design. For Economic Surveys, the most common type of sampling frame is a Business Register of all the establishments in the country, with information of the type of activity, number of persons engaged (employees) and any other measures of size from available sources, such as capital investment, revenue, etc. This information will make it possible to design an efficient sample for the Economic Surveys and, with this type of list frame, the sampling unit is the individual enterprise or establishment.

In the case of cluster sampling, the sampling frame will include all the primary sampling units (PSUs), such as segments in an area frame, that cover the units of analysis (establishments) in the population. In the case of an area sample, an updated listing of establishments in each selected segment will provide the sampling frame for the second stage of selection.

2.2   Simple Random Sampling

Simple Random Sampling (SRS) is the most basic type of sampling, where each sampling unit is selected randomly with equal probability. In practice, simple random sampling is generally without replacement (SRS-WOR), meaning that once a unit is selected it is removed from the frame before selecting the next unit so that an individual unit cannot be selected more than once. In the case of relatively high sampling rates (for example, greater than five per cent), this has an implication for the calculation of sampling errors, which will have a finite population correction factor. Given that all the units are selected with equal

---

[1] Readers are encouraged to refer to the classic sampling texts such as *Sampling Techniques* (William G. Cochran) or *Survey Sampling* (Leslie Kish) for a more detailed coverage of sampling theory and concepts.

probability under SRS, the units all have the same weights or expansion factors; this is referred to as a self-weighting sample.

Simple random sampling is important because it forms the basis of sampling theory. However, pure SRS is rarely used for Economic Surveys as it would not be efficient given the considerable variation in the size of establishments. In the case of stratified sampling (see 2.4 below), SRS may be used within each stratum, which is treated as an independent (separate) population. In practice, pure SRS is only used in the case of small homogeneous populations.

2.3   Systematic sampling

In systematic sampling, units in the frame are selected at fixed intervals based on the sampling fraction. In the case of equal-probability systematic sampling, the sampling interval is calculated as the inverse of the sampling fraction. For example, in the case of a five per cent sample, the interval would be 1/0.05=20. A random number would be selected between one and the interval, identifying the first unit selected. In this example, let us assume that the random start is 12. In this case the selected units would be 12, 32, 52, etc. If the order of the frame is random, systematic sampling can be used as a convenient way to obtain a simple random sample. However, it is possible to order the frame by certain criteria to provide a sample that is more efficient than a simple random sample, as shown in later examples. Moreover, systematic sampling can also be used with probability proportional to size.

2.4   Stratification

Stratification is one of the most important concepts in designing an efficient Economic Survey. Each stratum is a separate group of homogeneous (similar) units within the sampling frame, treated as an independent population for the purposes of sampling. It is possible to use a different sampling rate and sampling strategy in the individual strata. In the case of an Economic Survey, the enterprises or establishments in the sampling frame are generally stratified by economic activity and size, as described in a later section of these guidelines. Different levels of stratification may be used, including geographic domains (such as provinces or states).

2.5   Cluster sampling

A cluster is defined as an area segment or concentrated group of units of analysis (establishments); each cluster should have well-defined boundaries. Cluster sampling is generally used in multi-stage sampling; in this case, the clusters are the PSUs selected at the first stage. A sample design can include different levels of clustering, such as towns selected at the first stage and segments selected at the second stage.

Cluster sampling is commonly used in household surveys and sometimes in establishment surveys in order to reduce the cost of the fieldwork by concentrating the sample in a limited number of sample areas. This will increase the operational effectiveness of the data collection and make it easier to supervise the interviewers. A new listing of establishments in the sample areas will provide an updated frame for representing the newer establishments.

At the same time, cluster sampling generally increases the variance and sampling error of the survey estimates compared with a simple random sample. This is due to the homogeneity or similarity of establishments within a cluster. The design effect is defined as the variance of an estimate based on the actual sample design divided by the corresponding variance from a simple random sample of the same size; it is a measure of the relative efficiency of the sample design. In cluster samples, the design effect is generally higher than 1.0. On the other hand, a stratified one-stage sample of establishments normally results in a design effect much lower than 1.0, indicating it is more statistically efficient than a simple random sample.

An enterprise with multiple establishments can also be considered a cluster, even though the establishments may be geographically dispersed. In this case, it would be appropriate to use the enterprise with all its establishments as the sampling unit if the accounts and information on employees are maintained at the headquarters. Such large enterprises would generally be included in the sample with certainty so that they do not contribute to the sampling error of the survey estimates.

2.6   Sampling with Probability Proportional to Size

A common sampling procedure used for sample surveys is the selection of units with Probability Proportional to Size (PPS), generally combined with systematic sampling. This is most frequently used for selecting PSUs in multi-stage sampling. A measure of size is obtained for each sampling unit, used for determining the probability of selection. This improves the efficiency of the sample design by selecting the larger units with a higher

probability and reducing the effect of size variability on the sampling errors of survey estimates. The weighting procedures should take into account the different probability of selection for each unit. In the case of an area sample, the measure of size can be the number of establishments or the total number of persons engaged in the segment as recorded in the most recent Economic Census data. For a one-stage sample of establishments, the measure of size can be the number of persons engaged (based on information in the sampling frame). However, it is generally more common practice to use stratification by size, rather than PPS sampling of establishments, since the former is efficient for establishing a certainty stratum for the largest units, and a different sampling rate can be used for each smaller size stratum.

2.7   Sampling error

The Economic Survey sample estimates are subject to sampling errors, given that they are based on data for a particular sample of establishments and not on information from the entire population. If a different sample of establishments were selected, the value of the estimate would be different. Conceptually, the variance is the average squared difference between the estimates from all possible samples and their mean. The standard error, or square root of the variance, is used to measure the sampling error. Other measures of precision are the coefficient of variation (standard error divided by the value of the estimate), which is a measure of the relative sampling error, and the confidence interval, a range of values (defined by the lower and upper limits) that includes the true value of the parameter with a stated probability (for example, 95 per cent).

3.0   The Objectives of Economic Surveys

In order to monitor their economies, individual countries have established a series of economic indicators, generally based on international standards. The most common indicators at the national level are the sets of national accounts, including the gross domestic product (GDP). The United Nations (UN), International Monetary Fund (IMF), World Bank and other international organizations have established international standards for national accounts. The standard currently being used by many countries is the System of National Accounts 1993 (SNA93). Economic Surveys are the main source of annual economic information for most sectors needed for the national accounts. Some countries are developing quarterly national accounts, requiring more frequent and timely information from sample surveys.

The data from Economic Surveys are also used for national economic planning and policy development. It is important to have timely and accurate economic information to develop programmes for particular sectors, and for monitoring and evaluation to measure the impact of different policies and programmes.

For designing any type of survey, the best practice is to begin with developing the outlines of all the tables that will be produced from the survey data. In the case of Economic Surveys, the tables will indicate the types of variables that will be tabulated, the economic groups that will be treated as domains and the geographic levels of disaggregation. The content of the tables will also determine all the items that need to be included in the survey questionnaire.

The main objective of many annual Economic Surveys is to produce reliable results on total employment, gross output, intermediate consumption and value added for different activities within each economic sector. The individual economic activities are identified by the 4-digit International Standard Industrial Classification (ISIC) codes. The ISIC coding system is maintained by the Classifications Registry of the United Nations Statistical Division (UNSD). It was originally adopted in 1948 and has undergone various revisions. The latest ISIC revision 4 was adopted in 2006.[2] The major 2-digit ISIC divisions of industrial sectors are presented in Annex 1 of this chapter.

In order to determine the domains of analysis for the tabulations from the Economic Survey data, it is important to first tabulate the distribution of the total number of establishments, total employment and revenue at the national level for each 4-digit ISIC activity, if such data are available from a recent Economic Census or Business Register. When a country has few establishments within a particular 4-digit ISIC activity, it will be necessary to include all these establishments in the sample with certainty if separate survey estimates are required for this activity. If there are many such small 4-digit activities, the resulting large sample size could exceed the resources available for the survey. Sometimes, these smaller economic activities are combined to the 2- or 3-digit ISIC level. On the other hand, some 4-digit ISIC activities will have many establishments, so they can be treated as individual domains of analysis.

In line with the survey objectives, it is also necessary to define the geographic levels for the tables, such as region, state, province or governorate. One problem is that, if each economic activity at the 4-digit ISIC level is crossed by all the provinces in a country to establish the domains of analysis, there will be many categories with few establishments. In this case, it will be necessary to include all the establishments in these small categories in the survey with certainty. Again, this will make the sample size too large and the survey would be costly. The level of disaggregation in the survey tables will directly affect the

---

[2] See: http://unstats.un.org/unsd/cr/registry/regcst.asp?Cl=17.

sample size so, sometimes, it may be necessary to limit the geographic domains or group similar activities with few establishments. For example, in some countries, the domains of analysis are limited to the 2-digit ISIC level within each province. This is why it is practical to begin with a frequency distribution of the number of establishments by geographic area and different levels of ISIC groups. In this way, the categories can be combined as necessary.

In some countries individual Economic Surveys are conducted for the different sectors (manufacturing, retail and wholesale trade, services, etc.) with special staff responsible for developing and updating the frame of establishments for each sector. For example, some countries conduct a periodic Industrial Survey given the importance of manufacturing, quarrying and mining to their economy. In other countries, an integrated approach is used for an annual Economic Survey that covers most economic sectors.   As mentioned previously, agricultural and livestock production are generally covered by special household-based surveys.

Another common objective of Economic Surveys is the estimation of trends or changes in certain indicators over time. This is important for monitoring the overall direction of the economy on a regular basis. An overlap in the sample establishments from one reporting period to the next will improve the precision of the estimates of trends; this is especially important for the larger establishments, which are generally included in the sample with certainty (that is, with a probability of 1).

4.0   Sampling frames for Economic Surveys

The sampling frame for an Economic Survey is generally based on an Economic Census frame of establishments, a Business Register, or area units that cover the entire population defined for the survey. It is important that the sampling frame is updated over time to include the new (birth) establishments and to exclude any establishments that no longer exist (deaths). In some countries, a combination of list frame and area frame are used for covering different parts of the population of establishments.

Developing a good sampling frame of economic establishments is generally the most challenging part of planning and conducting Economic Surveys, given the dynamic nature of most economies. In view of the lack of resources to invest in updating the sampling frame, some countries base their Economic Surveys on an outdated directory of establishments and, thus, obtain biased results. This is fairly obvious when the frame of establishments becomes smaller over time, as some older sample establishments no longer exist while birth establishments are not added to the frame.

This section describes different types of sampling frames that can be considered for annual and quarterly Economic Surveys.

4.1   Economic Census

Many countries conduct an Economic Census of Establishments at least every ten years. In the case of more developed countries, a census is conducted every five years because of the dynamic nature of the economy. Many developing countries conduct a field operation for the Economic Census to enumerate all the economic establishments within enumeration areas (EAs). More developed countries, with good systems of registration and business directories, may depend more on a census carried out by mail.

The field operation for the Economic Census sometimes adapts the cartographic materials developed for the Census of Population and Housing. During the census cartographic operation, small EAs with well-defined boundaries are delineated. Each EA corresponds to the workload of one enumerator. Sometimes the establishments are listed as well as the households in each EA prior to the census. The Economic Census should cover all the establishments in each EA, whether they are registered or not. The Census questionnaire generally covers structural information to establish the size and distribution of economic activities which can later be used as a sampling frame for more comprehensive periodic sample surveys.   The questionnaire should include all the geographic and address information required to develop or update the Business Register, including the name of the establishment and all the most recent contact information such as telephone and mobile phone numbers and e-mail addresses. The form should include sufficient information on economic activity for coding the ISIC at the 4-digit level. The information should include the total number of persons engaged and other measures such as capital investment, annual value of production, or sales. Sometimes a shorter and simpler form is used for the smaller establishments.   It is also possible to use more specialised questionnaires for individual economic sectors.

The quality of the sampling frame depends on both the coverage of the Economic Census and the accuracy of the information on size and description of the economic activities. It is important to carefully control the quality of the ISIC coding operation in order to minimize the misclassification of economic activities.

The most recent Economic Census data can be used as the base for developing the sampling frame of establishments, which should be updated over time from various sources. It is important to match the Economic Census database to any previous Business Register and other available directories in order to evaluate the coverage of both the Economic Census frame and the Business Register and to update the frame of establishments using both

sources of data. A quality control operation should be conducted to ensure that all the active large establishments (for example, with 50 or more persons engaged) in the Business Register are included in the Economic Census frame. This matching process should begin with the largest establishments and continue down to the smaller establishments when resources are available. In addition to using the number of persons engaged as a measure of size, the establishments should also be ranked by size in terms of the value of production or sales, when this information is available. This procedure is similar to that used for updating the frame of establishments over time.

4.2   Business Registers

A Business Register or frame of economic enterprises and establishments that is updated over time is the best sampling frame that can be developed for Economic Surveys. It is important that this Register should include all the information needed for each establishment in order to develop a good sampling frame for Economic Surveys. This information should include a description or code for the economic activity, the name and address of the establishment, contact information (telephone and e-mail), the number of persons engaged, value of annual production or sales, amount of capital investment or other measures of size. It is ideal if the registration system includes annual licensing, where the information can be updated with the current number of persons engaged.

An important issue in developing an effective sampling frame for Economic Surveys is to establish a linkage between the enterprise and individual establishments of the enterprise. A hierarchical coding system of enterprises and establishments can be used for the purpose. This is especially important for the larger enterprises that may have many different establishments. It is also ideal for the frame to include information on the type of management and accounting used by the enterprise as this will determine the most effective procedures for data collection. For example, if centralised accounts and employment information are maintained by the headquarters of the enterprise, then the information for all the establishments can be obtained at this level. On the other hand, if the establishments within an enterprise operate independently, the information will have to be collected at the establishment level. Sometimes, the nature of the management and accounting system for a multi-establishment enterprise is included in the Business Register or Economic Census data.

Countries with a good enterprise or establishment registration system can use this as a basis for developing and updating the sampling frame. Such registration systems are sometimes maintained by the Ministry of Industry and Trade or another government agency for the purposes of issuing licences. In some countries, the Ministry of Finance or the government agency responsible for taxation maintains a directory of enterprises or establishments for

the purposes of collecting taxes. The Statistics Office should establish agreements with such agencies to share their directories of enterprises and establishments. However, the Statistics Office should not provide information to the other agencies in order to strictly maintain the confidentiality of the data collected from the establishments. The reputation of the Statistics Office in enforcing its confidentiality regulations is important for gaining the cooperation of the sample establishments.

Many countries also have unregistered or informal establishments that are an important part of the economy. In this case, an effort should also be made to include these establishments in the sampling frame.

4.3   Unique identification number for enterprises and establishments

In order to maintain a Business Register and establish a sampling frame of enterprises and establishments, it is ideal to have a unique permanent identification number for each enterprise and establishment that is also shared with other government agencies that are sources of information for the Business Register, such as the Ministry of Industry and Trade and the Ministry of Finance. Some countries use the tax identification number of this purpose. This will facilitate the matching of establishment directories from different sources in order to update the sampling frame of enterprises and establishments over time.

An inter-ministerial committee should be established to reach agreement on a system of common identification codes for establishments. A procedure for the timely inclusion of birth establishments in the frame with appropriate identification codes should be worked out. The records of establishments that no longer exist should be removed from the active frame and separated into historical files. In order to ensure the confidentiality of the data collected by the statistics office, it is critical that this organization does not share with other government agencies any micro data with personally identifiable information from the Economic Surveys. In cases where the Business Register includes both enterprises and individual establishments, it is ideal to have a coding system linking the enterprises and individual establishments.

4.4   Multi-stage sampling from list frame

Even with a list frame of establishments, some countries use multi-stage sampling procedures in order to reduce costs. If the Economic Survey data are collected through personal interviews, it can be costly to have the sample dispersed through all towns and rural areas of the country, both in terms of transport costs and staff time; supervision also becomes more problematic with a dispersed sample. In this case, primary sampling units

can be established based on the geographic hierarchy of the country. For example, some countries define towns or localities as PSUs that are selected at the first sampling stage. In this case, the sample establishments for the Economic Survey would only be selected from the list frame of establishments for the sample towns and localities. With this type of sampling scheme, the towns and localities would generally be selected with probability proportional to size, where the measure of size is based on the total number of establishments or persons engaged in the PSU. Using the number of persons engaged as the measure of size is preferable in order to increase the probability for towns with larger establishments. With a PPS selection, the largest cities will be included in the sample with certainty. It is recommended to identify all these self-representing cities and move them into a separate frame, with the sample establishments allocated proportionately to each city based on the total number of persons engaged.

When this type of multi-stage sample design is used, it is recommended to have an independent frame for the very large establishments (for example, with a minimum of 100 or 200 persons engaged), regardless of where they are located. The number of such establishments should be fairly small, so this would only slightly increase the transport costs for the survey. Most of these large establishments are probably located in the self-representing cities that are included in the survey anyway. Perhaps a special arrangement can be made with these very large establishments to obtain the survey data electronically or through the mail in order to reduce costs. These very large establishments would be removed from the data for the frame of PSUs in order to avoid duplication in the two independent frames.

If enterprises are used as secondary sampling units (SSUs) in a multi-stage sample design, it is important to take into account the location of the individual establishments within a sample enterprise. If the management information and accounts are centrally maintained at the enterprise headquarters, the best option would be to associate the establishments in all locations with the PSU where the headquarters of the enterprise is located. In this case, establishments in sample PSUs that have their headquarters in another PSU should be excluded from the second stage sampling frame. If the management information and accounts are kept by the individual establishments, the best option would be to include in the survey only those establishments that are located in the sample PSUs.

Once the self-representing and non self-representing sample PSUs (cities, towns and localities) are selected, the second stage selection of enterprises or establishments involves similar procedures to those described in the following sections of these guidelines (in terms of stratification by ISIC group and size, sample allocation, etc.).

4.5   Area sample as part of a multiple frame

Some countries use an area sample in order to complement the frame of establishments by covering the new birth establishments and those missing from the Business Register. Sometimes an area frame is also used to cover particular groups of smaller establishments that are numerous and geographically dispersed, as well as unregistered or informal economic activities.

 The main advantage of an area frame is that it provides a cost effective means of identifying and covering birth establishments for the Economic Survey. It is especially efficient for updating the frame for smaller birth establishments that are not captured in the Business Register. In this case, a sample of segments can be selected in one or two stages and a new listing of establishments can be conducted within the boundaries of these sample segments to provide a frame that represents all the birth establishments in the country. A complete enumeration of establishments is generally limited to an Economic Census conducted every five or ten years. Although an area sample is subject to higher sampling errors due to clustering effects, there is a significant cost reduction by updating the frame and conducting the survey interviews in a representative sample of areas.

Unlike the case of a multi-stage sample based on a list frame described previously, the size of the PSU or segment should be small enough to be listed by an enumerator. For this reason, it is convenient to use the EAs defined for the Economic Census as the PSUs when this type of frame is available. In this instance, the Economic Census can be used for developing a summary database with measures of size (total number of establishments, total persons engaged, value of production, etc.) depending on the information available. Such a database can then be used for selecting a sample of EAs with PPS. In some countries, more complex measures of size are used that also take into account the variety of economic activities in the EA. An area sample would be stratified by geographic and other criteria in order to improve the efficiency of the sample design.

When an area sample is used to represent new establishments and small establishments for some economic activities, the Economic Survey will be based on a multiple frame. The most important frame is still the Business Register or list of establishments, which is critical for providing complete coverage of all the larger establishments, as well as economic activities with few establishments in the frame. The large establishments are also easier to identify through the formal registration system and licensing programmes. The area frame would only be used to represent any establishments that were not included in the list frame, including new birth establishments. In this case, it would be necessary to match the establishments listed in sample areas to the Economic Census frame to ensure that there is no duplication. Sometimes, the area frame is also used instead of the Business Register to cover small establishments in particular economic activity groups that are numerous and

geographically dispersed. For these activities, the data collection for the Economic Survey should be concentrated in the sample areas in order to make the field operations more efficient.

The area frame in some countries is based on the frame of census enumeration areas (EAs) from the most recent Economic Census. In order to compile the area sampling frame, it is first necessary to produce a database with summary information by EA from the Economic Census data, such as the total number of establishments by sector or subsector. In order to increase the sampling efficiency, the sample EAs can be selected with PPS. The measure of size used depends on the sampling strategy. The simplest measure would be the total number of establishments, indicating the level of economic activity in the EA. This would be appropriate in the case of a multiple frame where the area frame is used to update the frame for new birth establishments. Alternatively, the total number of persons engaged in all establishments within the EA can be used as the measure of size to give higher weight to the EAs with larger establishments. In most countries, establishments with similar economic activities have a tendency to be concentrated or clustered together and this should be taken into account in the design of the area sample.

Sometimes an area sampling frame is used to cover all the smaller establishments for economic activities that are commonly found spread throughout the country, such as tailors, barber-shops and small general stores. If such small establishments are only to be covered by the area frame (and not the Business Register), the measure of size can be based on the total number of all small establishments (or number of persons engaged) covered by this frame in each EA from the last Economic Census.

Some countries use a more complex indicator such as the measure of size, based on the variety of economic activities as well as the total number of persons engaged. The methodologies used by two different countries are presented here to illustrate alternative applications of area sampling frames.

4.6   Area sample - example one

In this example, a country based the sampling frame for its annual Economic Survey on a list frame of establishments to cover the establishments with 10 or more persons engaged (from the Economic Census, with updates from different sources); and an area frame for the establishments with less than 10 persons engaged. The area frame may also provide the information for larger establishments for which data in the list frame might be missing.

The Statistics Office established 160 different ISIC groups as domains for the tabulation and analysis of the Economic Survey data, based on the most recent Economic Census data. Each group corresponded to 2, 3 or 4-digit ISIC activities, depending on the total number of establishments in the frame and the similarity of the economic activities. For example, if a 4-digit ISIC activity had many establishments, it could be treated as one group, whereas similar activities with few establishments were grouped together.

First, an EA summary file was developed with a count of the total number of persons engaged (only for establishments with less than 10 persons engaged) in each of these 160 ISIC groups by EA. The establishments with 10 or more persons engaged are excluded from this measure since they are covered in the list frame of establishments. This summary matrix, or database, had 160 fields or columns for the ISIC groups, and one record or row for each EA. The sum for each column corresponds to the total number of persons engaged in that ISIC group (in establishments with less than 10 persons engaged). Given their relatively small size, most EAs only have a few economic activities so most of the cells in this matrix are 0.

Next, a new matrix of the same dimension was generated by dividing the number of persons engaged in each EA within an ISIC group (from an individual cell) by the total for that group (column); this matrix represents the proportion of all persons engaged in each ISIC group belonging to that EA.

 Finally, the measure of size for each EA is calculated as the sum of the proportions for all the 160 ISIC groups (that is, the proportions are summed across each row). Since the sum of all the proportions within each column (individual ISIC group) will sum to 1, the total cumulative measure of size would be 160.

In this example, the measure of size for an EA increases based on the number of persons engaged in each economic activity within the EA as well as the total number of economic activity groups represented in the EA. This measures the relative importance of the EA in contributing to the estimates for the different ISIC domains in the case of establishments with less than 10 persons engaged.

The EAs were then sorted by measure of size to determine the cut-off in order to identify any EAs to be included in the sample with certainty. This also set a minimum measure of size thus avoiding extreme variations in weights and ensuring a minimum probability of selection for the smaller EAs.   This sampling scheme was actually implemented separately in eight different geographic strata (four regions, urban and rural). The sample EAs were allocated to these geographic strata in proportion to the cumulative size of the stratum, determined in a similar manner at the geographic stratum level.

Based on the objectives and budget for the annual Economic Survey, a sample of more than 2,000 EAs was selected. A new listing of all the establishments within each sample EA was conducted. The listing sheet included information to identify the economic activity and number of persons engaged in each establishment. Given the relatively small number of establishments in each ISIC group listed in most sample EAs, the general procedure was to include in the survey all the establishments with less than 10 persons engaged. However, if the number of establishments in a particular activity exceeded a certain number (for example, 10), the enumerator contacted the supervisor and a second stage sampling procedure was used to select a systematic sample of 10 establishments for that ISIC group in this case. This was generally limited to cases where a market or large service sector was found in the EA. In the case of establishments listed with 10 or more persons engaged, they were matched to the list frame of establishments to determine whether they were included in that frame. If they were found in the list frame, they were excluded from the area frame selection in order to avoid duplication in the two frames.

4.7   Area sample - example two

For the annual Economic Survey in another country, the Statistics Office first identified the following 2-digit ISIC groups in the industry, services and internal trade sectors that have small establishments spread throughout many city blocks and localities: 15, 18, 26, 28, 36, 50, 51, 52, 55, 74, 80, 85 and 93. The Office determined that it would be more cost effective to select the smaller establishments in these activities using a two-stage area sample design. In this case, the data collection for the annual and quarterly Economic Surveys would be concentrated in the sample areas in order to make the field operations more efficient.

Based on data from a recent Economic Census, the Statistics Office identified 32 major cities and localities where the economic activities were concentrated. In these cities, the primary sampling units (PSUs) were defined as individual blocks. For the remaining areas, the localities are defined as PSUs. At the second stage of sampling, the establishments within each sample block or locality were selected separately for each of the three major economic sectors.

In order to ensure a good dispersion of the area sample while still coping with operational constraints, the Statistics Office decided to select a sample of 400 sample blocks and 80 sample localities. A sample of about 10 establishments was selected within each sample area at the second stage, considering the potential intra-class correlation (and resulting clustering effect) as well as the total number of establishments allocated to this frame based on the resource constraints. This sample size should provide a good representative sample for the specified 2-digit ISIC groups included in this frame. Based on the experience from the first year of using the area sampling frame and the sampling errors resulting from the

survey data, it would be possible to adjust the sample size and allocation in future years, if necessary.

The sample blocks and localities within each province were selected with PPS, where the measure of size for each block and locality was determined by the percentage of revenue covered by all establishments in that block within each 2-digit activity, as well as by the number of different 2-digit ISIC activities represented in the block. This measure of size was calculated by adding the percentage of revenue for each 2-digit ISIC category across all 13 groups. This measure of size will increase the efficiency of the sampling frame by giving a higher probability of selection for blocks with more revenue and more economic activity groups represented among those covered by the area frame.

A listing of all the establishments was conducted in each sample block. For the second stage of selection, these establishments were divided into three economic sectors and a specific number of establishments were selected from each sector within the sample EA. A total sample of up to 10 establishments was selected for all three sectors within the EA.

5.0   Updating the Business Register

In order to ensure that the results of the Economic Survey represent the current economy, it is important to continually update the Business Register or sampling frame of establishments to cover all new birth establishments and remove establishments that no longer operate. This will require special staff responsible for checking different sources of establishment directories, including lists from the Ministry of Industry and Trade and other government agencies responsible for registering and licensing establishments, chambers of commerce and industry, professional associations, and commercial directories such as the Yellow Pages.

It is important for the Statistics Office to have an arrangement with the Ministry of Industry and Trade and other government agencies to send regular updates on any new enterprises or establishments that are registered. As indicated earlier, it would be ideal to use a common unique identification number, such as a tax or employer identification code to make it easier to update the Business Register on a regular basis.

The staff responsible for the Business Register should be provided with sufficient resources for updating and maintaining the frame. They should have an adequate number of telephone lines, good Internet communication and transportation facilities for the frame updating operation. When a telephone number or e-mail address is available for an enterprise, these means of communication can be very cost effective for obtaining some

types of information for updating the frame. Some companies also provide various items of information on their websites.

In the process of updating the sampling frame, priority should be given to the larger establishments, since they will have a greater effect on the results from the Economic Surveys. Since the largest establishments will be included in the sample with certainty, the sampling frame can be updated with the most recent information following each Survey.

6.0   Stratification of Sampling Frame

One of the most important features of an efficient sample design is the stratification of the sampling frame into homogeneous groups.   The sample selection is carried out independently within each stratum, although it is also desirable to order the establishments by certain criteria to provide further implicit stratification when systematic selection is used.

The nature of the stratification depends on the most important characteristics to be measured in the surveys, as well as on the domains of analysis. The sampling frame of establishments for the Economic Surveys will have at least two levels of stratification. The first level will be by the groups of economic activities that correspond to the domains of analysis, down to the 4-digit ISIC level depending on the objectives of the survey. The second level of stratification of establishments within each of these ISIC groups will be by the number of persons engaged, which is also correlated with value of production, revenue, expenditures and other aggregates. This high degree of stratification of establishments by ISIC and number of persons engaged will greatly improve the efficiency of the sample design and will make it possible to allocate the sample among the strata in an optimum manner.

The first step is to establish a cut-off for the number of persons engaged in order to determine the large establishments to be included in the sample with certainty. It is convenient to set an overall cut-off to establish the certainty strata for all the ISIC groups within each sector, such as 20, 50 or 100 or more persons engaged. This will depend on the distribution of the establishments by the number of persons engaged and will vary by the size of the country and the economy. It is also dependent on the budget for the survey as well as logistical considerations. A practical way to establish the overall cut-off for the number of persons engaged in order to determine the certainty stratum is to sort the establishments within each economic sector in a descending order by number of persons engaged. A spreadsheet can be used to calculate the cumulative number of persons engaged down this list and the cumulative percentage of total employment covered. The total number of establishments and the percentage of employment covered can then be examined for alternative cut-off points, such as 20, 50, 100, 150 and 200 persons engaged. It is important that the total number of certainty establishments does not exceed what can

be covered within the survey budget. Some countries use a cut-off as low as 10 persons engaged for an Industrial Survey to establish the certainty stratum.

Some countries also use value of production, capital investment or other variables to rank the establishments and determine the certainty strata. The distribution of establishments by such indicators is examined using a procedure similar to that described above for employment. In this case, an establishment is included with certainty if it exceeded either the employment or value of production cut-off point. In addition, some ISIC groups with few establishments may be included in the sample with certainty. This is discussed later in the section on sample size and allocation.

The employment size strata to be defined for the sampling frame may vary by country but it is common practice to use practical ranges to distinguish the different size classes, such as the following ranges of persons engaged:

- 1 person

- 2 to 4 persons

- 5 to 9 persons

- 10 to 19 persons

- 20 to 49 persons

- 50 to 99 persons

- 100 to 149 persons

- 200 or more persons

Depending on the cut-off for the employment size to determine the establishments to be included in the sample with certainty, several of these categories will become certainty strata. This will vary depending on the size of the country and its economy. In some countries, a cut-off as low as 10 persons engaged is used to define the certainty strata, while other countries may use a cut-off of 50 or 100 persons engaged. The cut-off for the certainty strata may also vary by economic sector.

If any of the employment size strata have a relatively small number of establishments, it can be combined with the previous or following category. For example, if a small number of establishments have 1 person engaged, this stratum can be combined with the next category to form a stratum of 1 to 4 persons engaged. However, the establishments in the combined strata can still be ordered by the original size group in order to provide implicit

stratification by size. For the non-certainty establishments in each ISIC group, a different sampling rate will be used for each employment size stratum, based on an approximately optimum allocation. However, for specific ISIC groups, it is also sometimes necessary to combine the size strata which have only a few establishments.

Within each ISIC group by employment size stratum, the establishments can be ordered by geographic and other criteria to provide further implicit stratification once the sample establishments are selected systematically within each stratum. This level of ordering should provide a very representative and efficient sample of establishments within each stratum.

In the case of an area frame, the PSUs are generally stratified geographically by state or province, urban and rural. It is also possible to classify the EAs by predominant sector, such as industrial, commercial and services.

7.0   Misclassification or change in classification of ISIC and size group

Given the dynamic nature of most economies, the number of persons engaged for a sample establishment may change during the time period between the Economic Census or the compilation of the Business Register and the Economic Survey. There may also be cases where the ISIC for a particular sample establishment has been misclassified; some establishments may change their primary economic activity; and the number of persons engaged may also change over time.

In these cases, it is important to maintain the original stratification codes of such establishments for the purpose of calculating the weights and sampling errors. The probability of selection of each sample establishment is based on its original stratification in the sampling frame. At the same time, the new ISIC code and current number of persons engaged should be entered in the questionnaire and used for the purpose of tabulating the survey data by the current categories.

The misclassification of ISIC group and employment size will increase the sampling errors of the survey estimates. Since the original stratification is used for the calculation of variances, the survey estimate for a particular ISIC group may have variance components from other ISIC strata.

When cases of misclassification and changes in sample establishments are identified, it is important to update the Business Register with the most current and correct information in order to improve the effectiveness of the next sample based on that frame.

8.0 Cut-off Sampling

Cut-off sampling involves including all the units (enterprises or establishments) above a certain size threshold in the survey with certainty and excluding the units below this size from the frame. The size is generally based on employment, value of production, or other information available in the sampling frame. This type of sampling is generally used for variables with skewed distributions where a relatively small proportion of the units in the frame account for a large proportion (for example, 80 or 90 per cent) of the total being estimated (persons engaged, value of production, revenue, etc.). Sometimes the value for the smaller units is estimated from previous data or modelling. In some cases, the largest units may be used to get a cost effective estimate of the trends or changes over time; for example, in small monthly surveys of the largest manufacturing establishments. However, some practitioners prefer to avoid this type of pure cut-off sampling, given that it is not a probability sample. This also depends on the objectives of the particular survey. For example, if there is also a need to study the characteristics of micro-enterprises, cut-off sampling would not be appropriate.

Another type of cut-off sampling involves dividing the frame into at least two strata. The units with a size above the cut-off are included with certainty and those below the cut-off are sampled. This alternative is more commonly accepted since it is a probability sample. It is based on the same methodology as stratified sampling and consequently is not treated separately in these guidelines. A third type of cut-off sampling method involves having a certainty stratum, a stratum that is sampled, and another cut-off for eliminating the smallest units from the frame.

Although there are studies of theoretically optimum thresholds for establishing the cut-off values, in general, a practical cut-off is chosen after sorting the establishments by size in descending order and cumulating the values for the key measure (persons engaged, value of production or revenues). The number of establishments above certain standard cut-off points is examined, such as when the cumulative measure reaches particular percentage points, such as 80, 90, 95 and 98 per cent of the total cumulative value. In some cases, there may be a natural break in the distribution of the measure that would be convenient to use. Some indicators for measuring trends are based on a fixed number of the largest establishments, such as the top 300 industrial establishments. [3]

---

[3] For those interested in a more theoretical perspective of cut-off sampling, a good reference is "A Framework for Cut-Off Sampling in Business Survey Design" (Marco Bee, Roberto Benedetti and Giussepe Espa). This file is available for download at the following website:
http://www.econo.economia.unitn.it/new/publicazioni/papers/9_07_bee.pdf

9.0 Model-Based Sampling

The focus of these guidelines is on design-based sampling, with classic stratification, clustering and known probabilities of selection. Some practitioners also use model-based sampling for establishment surveys, especially when there are extreme differences in the size and characteristics of the sampling units. Model-based sampling uses the relationship between available auxiliary variables and survey variables in terms of a model to establish selection and estimation schemes that minimize the measure of error for the distribution of the estimator over all possible replications of the population given the model. The objective of this sampling method is to minimize the mean square error (MSE) of the model-based estimates. The MSE is a measure of the total error of a survey estimate and is used to quantify the accuracy of the estimate. The MSE of a particular survey estimate $\hat{\theta}$ is defined as follows:

$$MSE\,(\hat{\theta}) = Variance\,(\hat{\theta}) + Bias^2(\hat{\theta})$$

The variance is due to the variable error and is generally measured by the square of the sampling error. The bias is related to systematic errors from non sampling sources, model specification and estimation. It is equal to the difference between the expected value of the estimate over all possible samples based on the particular model and the true value of the estimate.

Under classic design-based sampling, most of the estimators are designed to be unbiased but the variance component of the MSE can be large with smaller samples. On the other hand, model-based sampling generally provides estimates with a certain amount of bias, while attempting to minimize the variance component of the MSE. Therefore, the effectiveness of model-based sampling depends on the level of the bias from model specification and other sources.

Cut-off sampling is a type of model-based sampling procedure where the largest establishments are included in the sample with certainty and a model may be used to estimate the component of an indicator for the smaller establishments. Sometimes, there is an overlap between design-based and model-based sampling schemes. For example, a design-based sampling procedure for establishment surveys generally includes a certainty

stratum of the largest establishments, which would also be included in the sample under a model-based sampling scheme.[4]

10.0   Sample Size and Allocation

The sample size for a particular survey is determined by the accuracy required for the survey estimates for each domain, as well as by the resource and operational constraints.   The accuracy of the survey results depends on both the sampling error, which can be measured through variance estimation, and the non sampling error, which can be partly controlled through quality assurance procedures.   The sampling error is inversely proportional to the sample size.   On the other hand, the non sampling error may increase with the sample size, since it is more difficult to control the quality of a larger operation.   It is therefore important that the overall sample size be manageable for quality and operational control purposes.

Standard sampling texts include formulas for determining the sample size needed to obtain a specified level of precision for a particular survey estimate, generally based on simple random sampling.   In the case of complex samples, a design effect is applied to this sample size to take into account the effects of stratification and clustering.   In the case of an Economic Survey based on a one-stage stratified sample from a list frame, the design effects are generally much lower than 1.0.   The standard formulas for determining sample size are generally not appropriate for an Economic Survey since the sampling frame has large establishments included in the sample with certainty and size strata with different sampling rates.   The simple random sample formulas would only be useful for examining the sample size for a particular domain of analysis.   It may also be possible to use more complex sample size formulas based on stratified sampling and taking into account the certainty strata and differential sampling rates by size stratum.

In most countries, the overall sample size for an Economic Survey is limited to a certain maximum number of establishments based on the survey budget and other resource constraints.   This will also affect the size cut-off to determine the establishments to be included in the sample with certainty.   Frequently the sample size needed to meet the precision requirements for all the ISIC and geographic domains originally established for the survey will exceed the resource constraints so there will be a need to compromise in finalising the objectives of the survey, for example, by combining some of the smaller domains.   Small ISIC groups can be combined with similar economic activities and the level

---

[4] Additional information on model-based sampling can be found in the following references: *Model Assisted Survey Sampling* (C. E. Särndel, B. Swensson and J. Wretman); *Models in the Practice of Survey Sampling* (Graham Kalton, International Statistical Review, Vol. 51, No. 2, August 1983, pp. 175-188); and *Comparison of Purposive and Random Sampling Schemes for Estimating Capital Expenditure* (T. S. Karmel; Malti Jain, Journal of the American Statistical Association, Vol. 82, No. 397, March 1987, pp. 52-57).

of geographic disaggregation can be limited in order to reduce the overall sample size requirements. This may involve an iterative process until the sample size is consistent with the resource constraints.

The criteria for establishing the size cut-off for including the large establishments in the sample with certainty are described in the section on stratification. Once the certainty strata have been established, the total number of establishments to be included in the survey with certainty can be determined. The initial total maximum number of non-certainty establishments to be selected for the survey can then be calculated by subtracting the number of certainty establishments from the total sample size for the survey.

Next, it is important to use information from the sampling frame to tabulate total employment, value of production, revenue and any other key measure available in the frame for each ISIC domain, doing so separately for the certainty and non-certainty strata. This summary information can be copied into a spreadsheet, which can then be used for calculating the percentage of the total value for each variable covered by the certainty strata. For example, in the case of some ISIC groups, the certainty strata may cover 90 per cent or more of some totals, while, for other groups, the total employment and revenue may be concentrated in the non-certainty strata. In cases where the total values are concentrated in the certainty stratum, it may be possible to use a lower sampling rate for the non-certainty establishments for that ISIC group.

The next step will be to determine the differential sampling rates by employment size strata. The sampling rate should be higher for the strata with larger establishments and greater variability. One method used in some countries is to examine the distribution of the establishments by employment size category and then assign convenient sampling rates to the strata that increase with employment size. Table 1 shows an example of such a sampling scheme.

Table 1. Example of Sampling Rates by Employment Size Strata

	Employment Size Stratum (Number of Persons engaged)					
	1	2-4	5-9	10-19	20-49	50+
Sampling Rate	2.0%	5.0%	10.0%	25.0%	50.0%	100.0%

These sampling rates may provide a distribution of the sample similar to a Neyman optimum allocation, as discussed below. These rates would need to be adjusted to ensure that the overall sample size does not exceed the resource constraints. It is actually necessary to take

into account the distribution of establishments by size within each ISIC domain, since the requirements for precision are established at this level. It is also necessary to determine the percentage of employment (and other total values) that is covered by the certainty strata. For some activities, a particular employment size category may have few establishments, in which case it should be combined with the next size stratum for that ISIC domain. It is necessary to select a minimum of two establishments per size stratum. Within each ISIC domain, it is recommended to avoid giving a higher sampling rate to a smaller employment size group; it should also be combined with the next size stratum.

10.1 Example of sample allocation for national Industrial Survey

In order to illustrate this process, the actual sampling scheme for a national Industrial Survey in a developing country is described here. In this case, the domains were the individual 4-digit ISIC categories for manufacturing, construction, electricity, water, mining and quarrying. Given that establishments with 10 or more persons engaged cover a relatively large percentage of the total employment and industrial production, they were included in the sample with certainty. There were a total of 4,486 establishments with 10 or more persons engaged in the sampling frame, representing about 71 per cent of the employment in the frame. All of the establishments in the construction, electricity, water, mining and quarrying sectors are included in the survey regardless of size. Given that reliable estimates are required for each 4-digit ISIC group, additional activities with few establishments were also identified as certainty strata.

The sample allocation by employment size stratum was determined separately for each ISIC activity using an approximately optimum sample allocation. First, a Neyman optimum allocation by employment size stratum was determined for the non certainty establishments across all activities in order to provide a guide for the relative distribution of the sample by size stratum within each ISIC group. The following formula was used to calculate the approximate number of establishments to be allocated to each employment size stratum:

$$n_h = n \times \frac{N_h \times S_h}{\displaystyle\sum_h N_h \times S_h},$$

where:

n_h = number of sample establishments allocated to employment size stratum h

n = total sample size allocated to all non-certainty employment size strata

N_h = total number of establishments in employment size stratum h

S_h = standard deviation for number of persons engaged per establishment in size stratum h

The Neyman allocation assumes similar costs per interview in each size stratum. It can be seen that the optimum sample allocation to each stratum depends on the number of establishments in the sampling frame for that stratum, as well as the standard deviation for the key variable being measured in the survey. The number of persons engaged was used for calculating the standard deviation, since most of the survey estimates are correlated with employment. The employment data from the sampling frame were used to calculate the actual standard deviation for each size stratum across all ISIC activities. If data are available in the sampling frame for other variables, such as value of production or value added, such data could also be used for the Neyman allocation.

In this example, the overall sample size for the national Industrial Survey was about 6,000 establishments, based mostly on resource constraints. Subtracting the number of establishments to be included in the sample with certainty (4,880), we have a remaining sample (n) of 1,120 to be allocated to the non-certainty size strata. The Neyman optimum allocation was used to determine the overall approximate sampling rates needed for each employment size stratum. Table 2 shows the optimum sample allocation for the small establishments and the corresponding sampling rates by size stratum.

Table 2. Neyman Optimum Allocation of Sample Small Establishments

Parameter	Size Stratum (Number of Persons Engaged)			
	1	2-4	5-9	Total
N_h	3,263	11,331	8,060	22,654
S_h	0.26	0.79	1.28	
N_h x S_h	861.60	8975.95	10333.64	
Optimum n_h	48	498	574	1,120
Sampling Rate	1.5%	4.4%	7.1%	

In the case of the first size stratum with one person engaged, the actual standard deviation for the number of persons engaged is zero. However, taking into account the variability for other characteristics which will be included in the survey, the standard deviation for this size stratum appearing in Table 2 was arbitrarily set at one-third of the value of the standard deviation for the stratum of two to four persons engaged.

These overall optimum sampling rates by employment size stratum were used as a guide for determining the relative sampling rates for the size strata within each ISIC group.

In allocating the sample to individual ISIC by employment size strata, it was also necessary to ensure that each stratum has a minimum of two sample establishments. Given the small number of establishments in some of the strata, it was necessary to combine employment size strata for the corresponding ISIC groups. For example, a combined size stratum of one to four persons engaged was used for particular ISIC groups. In order to determine the number of sample establishments to be selected in each ISIC by employment size stratum and the corresponding sampling rate, it was also necessary to take into account the percentage of employment concentrated in the establishments with 10 or more persons engaged for each ISIC. For example, if the large establishments cover 90 per cent of the total number of persons engaged for the ISIC activity, the estimated total is less dependent on the sample of small establishments. In cases where the small establishments covered more than 50 per cent of employment for the ISIC activity, higher sampling rates were used.

Based on these various criteria, a preliminary sampling rate was determined for each ISIC by size stratum, which resulted in a total sample of more than 6,200 establishments (including the large establishments). Then, the sampling rates for the ISIC activities 1810 (manufacturing of wearing apparel by cutting and sewing) and 3611 (manufacturing of wood furniture) were adjusted downward since they have a very large number of establishments in the frame. The group of seamstresses and tailors is also probably more homogeneous than most activities. The adjusted sampling rates resulted in a total sample size of 6,004 establishments. After calculating standard errors for the survey estimates of total number of persons engaged by 4-digit ISIC, the sampling rates were increased for four ISIC groups with a coefficient of variation (CV) higher than 10 per cent. The final sample had a total of 6,020 industrial establishments. The CVs for the survey estimates of total number of persons engaged are now below 10 per cent for all ISIC activities, as shown in Annex 5. The distribution of the sample establishments by 4-digit ISIC activity and employment size stratum is illustrated in Annex 2, together with the corresponding sampling rates, sampling intervals and random starts used to select the sample establishments. The table also indicates the employment size strata that were combined for particular ISIC activities; in this case, the specified sampling rate is for the combined strata.

In the case of a multiple frame where an area sample is used to cover the smaller establishments for some economic activities, it will be necessary first to examine the distribution of the establishments covered by each frame, and estimate the percentage of employment, revenue or value added cover by each frame. The number of sample PSUs to select in the area frame depends on the level of resources as well as the distribution of the establishments in this frame.

11.0   Sample Selection Procedures

The sample selection methodology will be different for each type of sampling frame. Therefore the sampling procedures for each frame are described separately here.

11.1   Frame of establishments

In the case of the selection of establishments from a Business Register or sampling frame of establishments, it is convenient and effective to select the sample of establishments within each ISIC group by size stratum using random systematic sampling with equal probability. When the list frame of establishments is sorted by geographic and other criteria within each stratum, systematic sampling provides implicit stratification by such criteria and improves the representativeness of the sample. It is first necessary to separate the establishments to be included in the sample with certainty. Once the number of sample establishments to be selected in each stratum has been determined, the systematic sample is selected using the following procedures:

1) A serial number from 1 to N_h (the total number of establishments in the frame for the stratum) is assigned to the sorted list of establishments in each stratum.

2) The sampling interval for the stratum (I_h) is calculated as the inverse of the sampling rate, that is, N_h divided by the number of sample establishments to be selected (n_h): $I_h = N_h/n_h$

3) A random number (R_h) between 0.01 and I_h, with two decimal places, is generated as the random start.

4) The sample establishments selected within the stratum are identified by the following selection numbers:

$$S_{hi} = R_h + [I_h \times (i-1)]$$, rounded up,

where:

i = 1, 2, 3,..., n_h

The i-th selected industrial establishment in the stratum is the one with a serial number equal to S_{hi}.

In the example of the national Industrial Survey described previously, an Excel spreadsheet was used to calculate the sampling intervals and generate the random starts for all the non-certainty strata, as shown in Annex 2. A new Excel file was then produced for the 6,020 sample establishments selected for the survey, with all the information from the sampling frame for these establishments. This spreadsheet was also used for calculating the basic sampling weights.

11.2 Area sampling frame

As discussed previously, the area sample is generally based on a frame of EAs from the Economic Census or other segments that are small enough for the establishments to be listed by an enumerator. In order to select the sample with PPS, the EAs are assigned measures of size to indicate the level of economic activity in the Area; different types of measure of size were also described previously.

The sampling frame of EAs will generally be stratified geographically (for example, by region or province, urban/rural) in order to provide implicit stratification and to ensure that the sample is geographically representative. Once the measure of size has been determined for each EA, the sample will be selected with PPS, using the following procedures:

1) Cumulate the measures of size down the ordered list of EAs within the stratum. The final cumulated measure of size will be M_h.

2) To obtain the sampling interval for stratum h (I_h), divide M_h by the total number of EAs to be selected in stratum h (n_h) specified in the sample design report: $I_h = M_h/n_h$.

3) Select a random number (R_h) between 0 and I_h. The sample EAs in stratum h will be identified by the following selection numbers:

$$S_{hi} = R_h + [I_h \times (i-1)],$$ rounded up,

where:

i = 1, 2, ..., n_h

The i-th selected EA is the one with a cumulated measure of size closest to S_{hi} but not less than S_{hi}.

An example of this type of PPS selection of EAs is presented in Annex 3, where the measure of size is based on the total number of persons engaged in the in-scope establishments in the EA. The procedure would be the same with any other measure of size for the PPS selection of EAs or segments in the area frame. An Excel spreadsheet can be used for the PPS selection or another software package such as the Statistical Package for the Social Sciences (SPSS).

A listing of all the establishments within each selected EA should be made. The listing sheet should include information on the name, address, telephone numbers and other contact information for each establishment, as well as a description of the economic activity and the number of persons engaged. It is also useful to include a question on when the establishment began operating in order to determine whether it is new. The establishments in the listing will be matched to those in the Business Register or frame of establishments to identify those that are new or otherwise missing from the frame. Assuming that a limited number of such establishments are found in an EA, they can all be included in the sample for the Economic Survey to represent the birth establishments and others missing from the frame.

If the area sample is used to cover the small establishments in particular subsectors (generally limited to 2-digit ISIC groups), all the small establishments in these ISIC groups will be included in the second stage sampling frame. If the number of such establishments is limited, they can all be included with a probability of one at the second sampling stage. If the number of establishments in a particular sector exceeds a certain limit, a sample of establishments can be selected systematically with equal probability, using a procedure similar to that specified previously for the list frame sample within each stratum. In the case of the second example described in the section on the area sampling frame, the small establishments covered by the area frame were divided into three sectors, and a specific number of establishments were selected from each sector within the sample EA. A total of 10 establishments were selected for all sectors within the EA.

12.0 Considerations for Non-response and Sample Replacements

There are two types of non-response in survey data. The first type is item non-response, where a questionnaire is administered to a sample establishment but the response is missing for certain questions. In this case, there are two basic options that are used in practice. One is to have a no response category in the tabulations; this is mostly considered for qualitative types of data such as classification variables. The second option is to impute the missing value based on: (1) historical data (especially in the case of large establishments); (2) using a ratio to another variable in the same questionnaire, such as using value of production to estimate missing cost of inputs; (3) using the value for a similar sample establishment; or (4) using the average value of that variable for the corresponding stratum (ISIC and employment size).

The second type of non-response is referred to as unit non-response, or non-interview, where a selected enterprise or establishment cannot be interviewed because of a refusal, no respondent available, or another reason. In the case of sample non-interviews, it is very important to obtain information on the reason for the non-interview (refusal, respondent absent, not found, moved location, no longer exists, etc.), which is required for the estimation procedures. This type of non-response reduces the effective sample size and can result in considerable bias so it is important to establish effective procedures for handling non-interviews. The main options are discussed in this section.

Given the potential bias from non-interviews, the Statistics Office should make every effort to follow up such cases and reduce the non-response rate as much as possible. This is especially important for the large establishments. In the case of large industrial enterprises that dominate a particular economic activity, senior staff members may need to contact the enterprise personally in order to obtain cooperation. Unfortunately, the Economic Surveys in some countries suffer from unacceptably low response rates, which reduce the quality of their national accounts and other important indicators.

One alternative that can be considered for taking into account the non-interviews is to adjust the weights for the corresponding stratum; this procedure is discussed in the section on weighting procedures. Another option is to impute the information for the non-interview based on historical data (for example, if the sample establishment was included in the previous Economic Survey or Census); this procedure is recommended for large establishments that have a greater impact on the survey estimates.

Another alternative considered in some countries is to replace the non-interview establishment with a similar one in the frame. The main advantage of selecting replacements for non-interviews is that it maintains the effective sample size in order to ensure that the level of precision is not reduced. However, this type of replacement

procedure does not eliminate the non-response bias. One disadvantage of replacing non-interviews is that this procedure can often be used for convenience when a particular sample establishment is considered difficult to interview and can thus introduce additional bias. For this reason, some countries prefer to increase the original sample size based on the expected non-interview rate; for example, if they expect a 90 per cent response rate, they would increase the original sample size by 10 per cent.

If a Statistics Office decides to select replacements for non-interviews, the procedures should be very well controlled by the supervisors and every effort should be made to interview the original sample establishment before replacing it.

In selecting the sample replacement for each establishment non-interview, it is important to identify a similar establishment in the same ISIC by size stratum as the original sample establishment in order to reduce the non-response bias. The sampling frame of establishments is generally ordered by ISIC, size stratum and geographic codes and is thus an ideal source for identifying the sample replacement establishments for non-interviews. In order to replace a particular non-interview establishment, it is necessary first to find the non-interview establishment in the sorted sampling frame. The replacement establishment could then be identified as the one appearing before or after the non-interview establishment in the sampling frame sorted list, which is in the same ISIC and size stratum and, ideally, in the same locality. It is very important to identify the replacements with a special code and to document the identification of the original sample establishment and the reason for non-interview. The response rate should be calculated in reference to the original sample and the replacement rate should also be stated in the survey publications.

In the case of the large establishments and others included in the sample with certainty, it is not possible to select replacement establishments so it will be necessary to use imputation or weight adjustment to account for the non-interviews. In cases where it is not possible to complete the interview for a very large establishment, special procedures will be needed to impute key variables based on historical information from the Business Register, a previous survey or census, or another source, as stated previously.

13.0   Weighting Procedures

In order for the sample estimates from an Economic Survey to be representative of the population of enterprises and establishments, it is necessary to multiply the data by a sampling weight, or expansion factor. The basic weight for each sample establishment will be equal to the inverse of its probability of selection. Since all survey data will be processed by computer, it should be easy to attach a weight to each sample establishment record in the computer files and the tabulation programmes can weight the data automatically.

The probabilities and weights depend on the sample design so the weighting procedures here are described separately for the list frame and area frame sampling.

13.1   List frame sample weights

For the large establishments included in the Economic Survey sample with certainty, the probability of selection and corresponding basic weight would be equal to one, since these establishments are self-representing.   A stratified one-stage sample design is used for selecting the establishments in the non-certainty strata so the probability of selection (or sampling rate) is defined as follows:

$$p_h = \frac{n_h}{N_h},$$

where:

n_h = number of sample establishments in ISIC by employment size stratum h

N_h = total number of establishments in the frame for ISIC by employment size stratum h

The basic weight for the sample establishments in each stratum is the inverse of this probability of selection and can be expressed as follows:

$$W_h = \frac{N_h}{n_h}$$

It is important to adjust the basic weights to take into account the non-interview rate within each stratum, including the certainty strata. This will also depend on whether replacement procedures are used for non-interviews. The weights should be adjusted for non-interviews as follows:

$$W'_h = W_h \times \frac{n'_h}{n''_h}$$

where:

W'_h = adjusted weight for the establishments in stratum h

n_h' = number of valid sample establishments selected in stratum h (excluding those which no longer exist)

n_h'' = number of establishments with completed interviews in stratum h (including replacements)

As mentioned previously, for very large establishments (such as those with 50 or more persons engaged), it is recommended to impute the missing information from historical data instead of adjusting the weights. The weights should not be adjusted for non-interviews in cases where imputation is used for the missing data.

13.2 Area frame sample weights

The probabilities of selection for the sample establishments in an area sample will depend on the actual sample design used for a particular survey. Most area frame samples for Economic Surveys include two stages of selection. Therefore, the general example presented here is based on a stratified two-stage sample design. A sample of EAs is selected within each geographic stratum at the first stage and a new listing of all the establishments in the sample EAs is conducted. The listed establishments in each sample EA may then be stratified by sector or major economic activity and a sample of establishments is selected within each stratum at the second stage. For some strata, all the listed establishments should be included in the survey with certainty at the second sampling stage. The probability of selection can be expressed as follows:

$$p_{hij} = \frac{m_h \times S_{hi}}{\sum_{i \in h} S_{hi}} \times \frac{n_{hij}}{N_{hij}} \, ,$$

where:

p_{hij} = probability of selection for the sample establishments in the j-th economic activity group within the i-th sample segment (such as EA) in geographic stratum h

m_h = number of sample segments selected in geographic stratum h

S_{hi} = measure of size (based on the indicator established for the PPS selection) for the i-th sample area in geographic stratum h

n_{hij} = number of sample establishments selected in the j-th economic activity group within the i-th sample area in geographic stratum h

N_{hij} = total number of establishments in the j-th economic activity group within the i-th sample area in geographic stratum h

In cases where all the establishments for an economic activity are included in the sample at the second sampling stage, the second term of this probability would be one.

The basic weight is calculated as the inverse of this probability of selection, as follows:

$$W_{hij} = \frac{\sum_{i \in h} S_{hi}}{m_h \times S_{hi}} \times \frac{N_{hij}}{n_{hij}}$$

A non-interview adjustment factor similar to that used for the list frame sample weight can be applied to this weight, ideally within the economic activity group in the EA. However, it may be necessary to aggregate the weighting cells if the number of sample establishments is too small or the non-response rate within the EA is high.

14.0 Types of Survey Estimates

The most common survey estimates to be calculated from the Economic Surveys are in the form of totals and ratios. The survey estimate of a weighted total can be expressed as follows:

$$\hat{Y} = \sum_h \sum_i W'_h \times y_{hi}$$

,

where:

y_{hi} = value of variable y for the *i*-th sample establishment in stratum *h*

The survey estimate of a ratio is defined as follows:

$$\hat{R} = \frac{\hat{Y}}{\hat{X}}$$

,

where $\hat{Y}$ and $\hat{X}$ are estimates of totals for variables *y* and *x*.

The survey estimates of means and proportions are special types of ratios. In the case of the mean, the variable *x*, in the denominator of the ratio, is defined to equal 1 for each sample establishment so that the denominator is the sum of the weights. In the case of a proportion, the variable *x* in the denominator is also defined to equal 1 for all elements; the variable *y* in the numerator is binomial and is defined to equal either 0 or 1, depending on the absence or presence, respectively, of a specified characteristic for each sample establishment.

When accurate auxiliary data for a certain variable are available from an independent source such as a recent Economic Census, such data can sometimes be used for making a ratio estimate of the total, which generally has a lower variance than a corresponding direct estimate of the total from the weighted survey data. This type of ratio estimation is also used for measuring trends (changes over time). In order to illustrate this type of ratio estimation, let us assume that data on the value of production are available in the Economic Census frame for each sample establishment in the Economic Survey conducted two years after the Census. In this case, the ratio estimate of the current total value of production can be calculated as follows:

$$\hat{Y}_R = \frac{\sum_h \sum_i W'_h \times y_{hi}}{\sum_h \sum_i W'_h \times x_{hi}} \times X$$

where:

y_{hi} = value of production for the *i*-th sample establishment in stratum *h* from the Economic Survey data

x_{hi} = value of production for the *i*-th sample establishment in stratum *h* from the previous Economic Census data

X = total value of production for all in-scope establishments in the frame from the previous Economic Census data

The ratio in this expression can be interpreted as the factor of growth (or decline) in the value of production in the two years following the Economic Census. For example, if this factor is 1.2, it indicates that the value of production increased by 20 per cent since the time of the Economic Census. Such ratios using data for the same sample establishments from two different annual surveys can also be used to obtain more reliable estimates of trends. The correlation of the numerator and the denominator of this ratio will reduce the variance (and sampling error) of the estimate. Ratio estimates have an inherent bias, which is generally relatively small compared with the standard error. Therefore, when quality data are available from an independent source, the ratio estimate of a total can be more accurate than a corresponding direct survey estimate of the total (that is, it has a lower mean square error).

15.0. Calculation of Sampling Errors

In the publication of the results from the Economic Survey, it is important to include a statement on the accuracy of the survey data. In addition to presenting tables with calculated sampling errors and confidence intervals for the most important survey estimates, the different sources of non sampling error should be described.

The standard error, or square root of the variance, is used to measure the sampling error although it may also include a small part of the non sampling error. The variance estimator should take into account the different aspects of the sample design, such as the

stratification and clustering. In order to avoid the time and effort required to develop custom variance programmes, an ideal solution is to use an available software package to tabulate the variances. Two software programmes available for calculating the variances for survey estimates from stratified sample designs such as those for the Economic Surveys are CENVAR, a component of the Integrated Microcomputer Processing System (IMPS), developed by the U.S. Census Bureau, and the statistical software package Stata. The IMPS and CENVAR software and corresponding manuals can be downloaded free of charge from the U.S. Census Bureau website[5]. CENVAR is menu-driven and user-friendly. It uses the data dictionary defined in the DATADICT component of IMPS. It can be used to calculate the variances of totals, means, proportions and other ratios. It produces subpopulation estimates for each category of a classification variable and these variables can be cross-classified. For each estimate, CENVAR calculates the standard error, coefficient of variation (CV), 95 per cent confidence interval and the design effect (DEFF). Both the CENVAR and Stata software packages use an ultimate cluster variance estimator.

In order to tabulate estimates of standard errors using CENVAR, it is generally necessary to produce a new data input file from the original survey data, in an ASCII (text) format. Since the CENVAR package will only accept one record type, it is necessary to generate one record for each unit of analysis in the CENVAR data input file. For example, in the case of the estimates by establishment, the CENVAR input file should have one record for each in-scope sample establishment. Each record in the CENVAR or Stata data input files should include fields for the stratum and cluster (PSU) codes, the weight and first-stage sampling rate, in addition to the classification and analysis variables that are required for the particular CENVAR analysis. The classification variables are used to produce subpopulation estimates for all their respective categories. The analysis variables are generally continuous variables, such as the number of persons engaged or value of production, or count variables, which are equal to one if the unit has a certain characteristic and to zero otherwise. CENVAR automatically creates a count variable named INTERCEPT, which is equal to one for each record. The INTERCEPT variable can be used to obtain the estimate of the weighted total number of units (for example, the total number of establishments), or it can be used in the denominator of a ratio in order to obtain a mean or proportion.

CENVAR does not accept any blanks in the data file. In the case of classification variables, any record with a blank should be imputed with a special code to identify "missing" or "not applicable." The CENVAR output will include estimates for these categories, which can be deleted from the tables that will be published.

The formula for the variance of the survey estimate of a total used by CENVAR and Stata can be expressed as follows:

---

[5] www.census.gov

$$V(\hat{Y}) = \sum_{h=1}^{L} \left[(1 - f_h) \times \frac{n_h}{n_h - 1} \sum_{i=1}^{n_h} \left(\hat{Y}_{hi} - \frac{\hat{Y}_h}{n_h} \right)^2 \right],$$

where:

$V(\hat{Y})$ = variance estimator of a total

$\hat{Y}_{hi} = W'_h \, y_{hi}$ = weighted value of variable y for the i-th sample establishment in stratum h

$\hat{Y}_h = \sum_{i=1}^{n_h} \hat{Y}_{hi}$ = weighted total of variable y for stratum h

$f_h = \dfrac{n_h}{N_h}$ = sampling rate for stratum h

L = number of strata

In the case of a stratified one-stage sample design such as the sample based on a list frame of establishments, this generalised formula for the estimate of the variance of a total can be simplified as follows:

$$V\left(\hat{Y}\right) = \sum_{h=1}^{L} \left\{ (1 - f_h) \times \frac{n_h}{n_h - 1} \times \sum_{i=1}^{n_h} \left[\frac{N_h}{n_h} \times y_{hi} - \frac{\left(\frac{N_h}{n_h} \times \sum_{i=1}^{n_h} y_{hi} \right)}{n_h} \right]^2 \right\} =$$

$$= \sum_{h=1}^{L} \left\{ (1 - f_h) \times \frac{n_h}{n_h - 1} \times \left(\frac{N_h}{n_h} \right)^2 \times \sum_{i=1}^{n_h} \left[y_{hi} - \frac{\sum_{i=1}^{n_h} y_{hi}}{n_h} \right]^2 \right\} =$$

$$= \sum_{h=1}^{L} \left\{ \left(1 - f_h\right) \times N_h^{\,2} \times \sum_{i=1}^{n_h} \left[\frac{\left(y_{hi} - \bar{y}_h\right)^2}{n_h \times \left(n_h - 1\right)} \right] \right\} =$$

$$= \sum_{h=1}^{L} \left(1 - f_h\right) \times N_h^{\,2} \times \mathrm{var}\left(\bar{y}_h\right) = s^2\left(\hat{Y}_{st}\right)$$

It can be seen that the last expression in this equation is the variance estimator for a total based on one-stage stratified random sampling (without replacement). In the CENVAR or Stata application, the cluster is the individual establishment. A coding system has to be established to uniquely identify each ISIC group by employment size stratum.

The expression $(1 - f_h)$ is the finite population correction factor based on the sampling rate for stratum h. In the case of the certainty strata, since $n_h = N_h$, the sampling rate f_h is equal to one, so the corresponding finite population correction factor is equal to zero. As a result, the variance component for the certainty strata is correctly calculated as zero. It is important to include the data from the certainty strata in the data file in order to obtain the combined estimates from all strata. The certainty strata generally contribute a large amount to the total estimate but have a zero variance thus lowering the coefficients of variation and the design effect for the combined estimate from all strata.

In the case of an area frame, the PSUs would be the sample areas such as EAs, which would be identified as the clusters in the CENVAR or Stata application. Given that each stratum is independent, it is possible to include the data from both the list frame and area frame in the same data file, with appropriate stratum and cluster codes to define the sample design for each frame.

The variance estimator of a ratio used by CENVAR can be expressed as follows:

$$V(\hat{R}) = \frac{1}{\hat{X}^2} \left[V(\hat{Y}) + \hat{R}^2 V(\hat{X}) - 2\,\hat{R}\,COV(\hat{X}, \hat{Y}) \right],$$

where:

$V(\hat{R})$ = Variance estimator of a ratio

$$COV(\hat{X},\hat{Y}) = \sum_{h=1}^{L} \left[(1 - f_h) \times \frac{n_h}{n_h - 1} \sum_{i=1}^{n_h} \left(\hat{X}_{hi} - \frac{\hat{X}_h}{n_h} \right) \left(\hat{Y}_{hi} - \frac{\hat{Y}_h}{n_h} \right) \right]$$

$V(\hat{Y})$ and $V(\hat{X})$ are calculated according to the formula for the variance of a total.

15.1. Example of CENVAR Application for Industrial Survey

The CENVAR application for tabulating the standard errors from the frame for the national Industrial Survey data described earlier is presented here as an example.   In order to estimate the approximate level of precision that can be expected from the sample of 6,020 establishments selected for the national Industrial Survey, the CENVAR software was used to calculate the standard errors, CVs and 95 per cent confidence intervals for the sample estimates of the total number of persons engaged by domain.   The National Industrial Census data on the number of persons engaged for the 6,020 sample establishments were used for this study.

A new data input file in a text format was generated for this CENVAR analysis, with one record for each of the 6,020 sample industrial establishments (including those included in the sample with certainty).   First, it was necessary to define the sample design variables in the data file.   For the non-certainty strata of small establishments, the following 5-digit stratum code was defined:

STRATUM = 4-digit ISIC (primary activity) + employment size stratum

The following codes were used for the size strata:

1 = 1 person engaged

2 = 2 to 4 persons engaged

3 = 5 to 9 persons engaged

6 = 1 to 4 persons engaged (for combined stratum)

7 = 1 to 9 persons engaged (for combined stratum)

The employment size stratum codes 6 and 7 were created for the ISIC groups where it was necessary to combine the original size strata for the sample selection.

In the case of the establishments included in the sample with certainty, the variance component is zero, so it is not necessary to identify the individual certainty strata in the file. The following stratum code was used for all the certainty strata: 10000.

For the CENVAR analysis, it is necessary to define a design variable for the primary sampling unit (PSU), referred to as the CLUSTER. Since a one-stage sample design was used for the national Industrial Survey, the PSUs are the individual establishments. In order to identify the PSUs uniquely within each stratum, the original 6-digit reference number of each establishment was used.

The basic weight for the establishments in each stratum was calculated with four decimal places. These weights were multiplied by 10,000 to eliminate the decimal point in the data file since the IMPS data dictionary specifies four implied decimals for the corresponding field. The last design variable is the sampling rate for each stratum, which also has four implied decimals in the data file. The first stage sampling rate for each stratum is used for calculating the finite population correction factor.

The following classification variables were included in the CENVAR data file: region, type of industry, major economic activity, and primary activity (4-digit ISIC). The only analysis variable included in this CENVAR data file is the number of persons engaged for each sample establishment. The CENVAR analysis produced estimates of the total number of persons engaged and corresponding measures of precision for each category of the classification variables.

The IMPS data dictionary for the CENVAR input file is shown in Annex D and the results from this CENVAR analysis are presented in Annex E. It can be seen that the CV for the sample estimate of the total number of persons engaged at the national level is only 0.24 per cent. All the CVs are below 10 per cent at the 4-digit ISIC level. As expected, the standard errors are 0 for the ISIC groups included in the sample with certainty. The design effects are all very low (mostly less than 0.1), indicating a very high sampling efficiency due to the high degree of stratification, the approximately optimum sample allocation, and the certainty strata of large establishments.

Annex 1:   ISIC Revision 4, Major Economic Subsectors (2-digit ISIC)

Code	Description
B - Mining and quarrying	
05	Mining of coal and lignite
06	Extraction of crude petroleum and natural gas
07	Mining of metal ores
08	Other mining and quarrying
09	Mining support service activities
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
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
D - Electricity, gas, steam and air conditioning supply	
35	Electricity, gas, steam and air conditioning supply
E - Water supply; sewerage, waste management and remediation activities	
36	Water collection, treatment and supply
37	Sewerage
38	Waste collection, treatment and disposal activities; materials recovery
39	Remediation activities and other waste management services

Annex 2:   Example of Sampling Rates, Sampling Intervals (S.I.) and Random Starts (R.S.) for Frame of Small Industrial Establishments by 4-Digit ISIC and Employment Size Stratum, and Distribution of Sample Establishments

Primary Activity (ISIC)	Strata by Number of Persons Engaged															10+ P.E.	Total No. of Sample Estab.	% Persons Engaged 10+ Stratum
	1 Person Engaged					2-4 Persons Engaged					5-9 Persons Engaged							
	No. Estab. Frame	Sampling Rate	No. of Sample Estab.	S.I.	R.S.	No. Estab. Frame	Sampling Rate	No. of Sample Estab.	S.I.	R.S.	No. Estab. Frame	Sampling Rate	No. of Sample Estab.	S.I.	R.S.	No. Estab. Frame		
1321	-	-	0			8	100.00%	8			12	100.00%	12			30	50	99.3%
1322	-	-	0			-	-	0			-	-	0			2	2	100.0%
1323	-	-	0			-	-	0			-	-	0			3	3	100.0%
1410	2	100.00%	2			11	100.00%	11			17	100.00%	17			43	73	92.3%
1422	2	100.00%	2			1	100.00%	1			9	100.00%	9			28	40	97.7%
1429	1	100.00%	1			1	100.00%	1			1	100.00%	1			1	4	98.6%
1511	-	-	0			7	Combined				8	13.33%	2	7.50	3.83	28	30	94.2%
1512	38	Combined				156	4.12%	8	24.25	6.80	70	7.14%	5	14.00	9.64	110	123	88.3%
1513	6	Combined				10	Combined				4	10.00%	2	10.00	8.58	11	13	95.3%
1514	54	3.70%	2	27.00	9.61	144	4.17%	6	24.00	14.84	94	7.45%	7	13.43	9.83	294	309	93.5%
1520	-	-	0			-	-	0			1	100.00%	1			7	8	99.4%
1531	250	2.00%	5	50.00	2.91	957	4.18%	40	23.93	6.17	128	7.03%	9	14.22	7.80	59	113	47.4%
1532	25	Combined				40	Combined				9	6.76%	5	14.80	6.37	9	14	76.6%
1533	2	Combined				3	Combined				4	22.22%	2	4.50	1.24	6	8	87.7%
1541	86	2.33%	2	43.00	16.78	299	4.01%	12	24.92	4.83	207	7.25%	15	13.80	10.73	125	154	63.2%
1542	-	-	0			1	100.00%	1			-	-	0			1	2	87.5%
1543	-	-	0			-	-	0			1	100.00%	1			7	8	99.6%
1544	-	-	0			1	100.00%	1			-	-	0				1	0.0%
1549	37	Combined				39	5.26%	4	19.00	5.90	42	7.14%	3	14.00	10.29	57	64	80.4%
1551	10	Combined				74	4.76%	4	21.00	9.71	26	7.69%	2	13.00	4.44	93	99	94.9%
1552	-	-	0			1	100.00%	1			-	-	0				1	0.0%
1553	35	Combined				298	4.50%	15	22.20	2.29	182	7.69%	14	13.00	10.60	35	64	45.6%
1554	-	-	0			43	4.65%	2	21.50	12.46	80	7.50%	6	13.33	1.46	54	62	79.8%
1600	-	-	0			-	-	0			-	-	0			2	2	100.0%
1711	57	3.51%	2	28.50	14.16	89	4.49%	4	22.25	19.25	138	7.25%	10	13.80	12.87	94	110	72.0%
1712	37	Combined				57	5.32%	5	18.80	11.28	62	8.06%	5	12.40	9.41	36	46	82.0%

Annex 3: Example of Selection of Enumeration Areas with Probability Proportional to Size (total number of persons engaged) for Area Sample

Total Cumulated Measure of Size =   4958
Number of Sample EDs Selected in Stratum =   8
Sampling Interval =   619.75
Random Start =   389.2411

Selection of EAs:

Serial Number	Systematic Selection	Identification No. Sample Blocks
1	389.24	390
2	1008.99	1009
3	1628.74	1629
4	2248.49	2249
5	2868.24	2869
6	3487.99	3488
7	4107.74	4108
8	4727.49	4728

Geographic Code	Block Code	Total No. Establishments	Total No. Persons Engaged	Cumulated Persons Engaged	Selection
2302	1	44	274	274	
2303	1	58	167	441	1
2303	2	56	324	765	
2305	1	60	124	889	
2306	1	40	122	1011	2
2307	1	33	87	1098	
2307	2	33	80	1178	
2308	1	42	228	1406	
2308	2	34	42	1448	
2309	1	20	60	1508	
2309	2	27	62	1570	
2310	1	39	83	1653	3
2310	2	39	226	1879	
2311	1	22	127	2006	
2312	1	27	133	2139	
2312	2	29	169	2308	4
2312	3	28	29	2337	
2312	4	26	121	2458	
2313	1	30	144	2602	
2313	2	29	98	2700	
2313	3	27	119	2819	
2313	4	36	246	3065	5
2313	5	24	98	3163	
2314	1	27	43	3206	
2315	1	22	137	3343	
2315	2	30	132	3475	
2315	3	30	97	3572	6
2315	4	29	127	3699	
2315	5	26	50	3749	
2315	6	30	157	3906	
2401	1	28	71	3977	
2401	2	18	41	4018	
2401	3	24	75	4093	
2401	4	27	141	4234	7
2401	5	29	156	4390	
2401	6	45	210	4600	
2402	1	38	97	4697	
2403	1	32	68	4765	8
2403	2	30	63	4828	
2404	1	34	130	4958	

Annex 4:   IMPS Data Dictionary for CENVAR Input File Used for Analysis of Standard Errors from National Industrial Survey Sample

```
Page    1                  Data Dictionary: SAMPEST           IMPS Version 3.1
                           Created: 06/12/04  20:43:21

                           Record Length:      35

    The following records have been defined:

    Record                 Record Type Value                   Max.
    Name                       (RECTYPE)         Required     Records
    ------                 ----------------      --------     -------
    ESTABLISHMENT                                    Y            99

The following COMMON items have been defined.  They occur on all records.
-----------------------------------------------------------------------------
Item (occurs)           Data            Item
  Subitem (occurs)      Type  Position  Len.      Value Name      Values
-----------------------------------------------------------------------------

STRATUM                  N        1-5     5

REFNUM                   N        6-11    6
```

Record Name: ESTABLISHMENT   Record Type:

Item (occurs)   Data   Item
  Subitem (occurs)   Type Position Len. Dec. Value Name   Values
----

Item (occurs) Subitem (occurs)	Data Type	Position	Item Len.	Dec.	Value Name	Values
WEIGHT	N	12-17	6	4		
RATE	N	18-22	5	4		
REGION	N	23-24	2	0		
INDUSTRYTYPE	N	25	1	0	Manufacturing	1
					Mining&Quarrying	2
					Elect.,Water,Gas	3
					Construction	4
MAJORECONACTIV	N	26-27	2	0	01	01
					02	02
					03	03
					04	04
					05	05
					06	06
					07	07
					08	08
					09	09
					10	10
					11	11
					12	12
					13	13
					14	14
PRIMARYACTIVITY	N	28-31	4	0	1321	1321
					1322	1322
					1323	1323
					1410	1410
					1422	1422
					1429	1429
					1511	1511
					1512	1512
					1513	1513
					1514	1514
					1520	1520
					1531	1531
					1532	1532
					1533	1533
					1541	1541
					1542	1542
					1543	1543
					1544	1544
					1549	1549
					1551	1551
					1552	1552
					1553	1553
					1554	1554
					1600	1600
					1711	1711
					1712	1712
					......	......
PERSONSENGAGED	N	32-35	4	0		

Annex 5: Example of CENVAR Application: Tabulation of Standard Errors for Estimates of Total Number of Persons Engaged, by Domain

Sample Estimates of Total Number of Persons Engaged by Domain

Category	Estimate	Standard Error	C.V. (%)	95% Confidence Interval Lower	95% Confidence Interval Upper	Design Effect	Number of Observations
NATIONAL	296,800	715	0.24	295,399	298,201	0.00	6,020
TYPE OF INDUSTRY							
Manufacturing	249,899	715	0.29	248,498	251,300	0.00	5,118
Mining & Quarrying	20,564	0	0.00	20,564	20,564	0.00	172
Elect., Water, Gas	12,453	0	0.00	12,453	12,453	0.00	224
Construction	13,884	0	0.00	13,884	13,884	0.00	506
MAJOR ECONOMIC ACTIVITY							
1-Mining & Quarrying	18,802	0	0.00	18,802	18,802	0.00	99
2-Electricity	1,762	0	0.00	1,762	1,762	0.00	73
3-Water	4,925	0	0.00	4,925	4,925	0.00	97
4-Construction	13,884	0	0.00	13,884	13,884	0.00	506
5-Food	60,186	235	0.39	59,726	60,646	0.00	1,073
6-Textiles	9,278	87	0.94	9,107	9,450	0.00	202
7-Clothing	58,036	518	0.89	57,021	59,051	0.01	1,284
8-Footwear	5,457	94	1.73	5,272	5,642	0.00	122
9-Sawmills, Wood Prod.	35,499	133	0.37	35,239	35,759	0.00	371
10-Printing	5,655	106	1.88	5,447	5,864	0.00	132
11-Concrete Products	5,519	115	2.09	5,293	5,746	0.01	136
12-Metal Products	13,824	187	1.36	13,457	14,191	0.02	423
13-Furniture	27,682	287	1.04	27,119	28,245	0.02	725
14-Other Manufacturing	36,291	108	0.30	36,079	36,503	0.00	777

Chapter 4: A Manual of Data Processing in Industrial Surveys

1.0 Introduction

Industrial statistics can be sub-divided into three types. First, there are the industrial (or economic) censuses conducted infrequently to establish the size and structure of the industry. Second, there are annual (structural) surveys that cover a wide range of information for an entire year, such as production, material inputs, labour inputs and costs, balance sheets, environmental impacts and, sometimes, also other financial data. Third, there are short-term indicators providing time series of output, inputs, or other characteristics of a representative sample of enterprises on a monthly or quarterly basis.

The International Recommendations for Industrial Statistics, 2008 (IRIS) and the present Guidelines provide the basic methodology, the classifications and data items to be collected and include draft sample questionnaires for the compilation of these different statistics. They also integrate them into the broader framework of economic statistics. Surveys of different kinds underlay these three types of industrial statistics.

For the censuses, two main data sets are normally compiled, a Business Register and a set of characteristics for all industrial enterprises. The former may be available from external sources but, in developing countries, it is normally built up from a comprehensive survey as the first part of the census. The second set of data collected for an industrial census consists of the detailed output, input, labour, environmental and financial information on the enterprises and their establishments in the reference year. The data contents and survey design should be the same as that used for annual surveys so that no problems arise when comparing the census information with the annual data collected in subsequent years (the census replaces the annual survey in the census year). The main difference between census and annual enumerations is that the census covers a much larger number of units, while annual surveys are restricted to the larger units above a certain cut-off point.

Monthly or quarterly indicators are collected from a sub-sample of the units covered in the annual surveys and are often expressed as indices. This chapter deals with data processing issues relating to the surveys for Business Registers and annual Surveys only. The target audience here is the ICT (Information and Communication Technology) staff in National Statistics Offices. From a data processing point of view, most of the pertinent issues involved in Industrial Surveys are similar to those arising in other surveys typically carried out by a Statistics Office. This chapter presents the principles and illustrates them with practical examples.

The second section shows the typical steps involved in the life cycle of an Industrial Survey data processing system and defines the differing roles of the data processing and statistical

staff. It then outlines the broad issues involved from an automation point of view, raising and referencing them again in the context of examples given in the following sections. Practical examples presented in the remaining sections share experiences based on work done with national accounts and Statistics Offices in a number of countries (the Maldives, Ghana, Saudi Arabia, Lebanon, Malawi, and Iraq) since 2000.

The third section identifies some of the more important factors to consider when designing and building data processing solutions for surveys. Also presented are various options used for databases and programming for survey processing and an indication of the reasons for choosing one over the other. One of the important conclusions from this section is the need to equip automation staff with generic programming solutions which they can then customize.

In the fourth section, a generic survey-based reporting module is introduced. This module can be used to prepare tabulation reports for most Industrial Surveys. Customising the programme for any tabulation entails specification of an Extensible Markup Lanugage (XML) file, which contains the database and survey- related details.

In the fifth section, an introduction to PDA (portable digital assistant or hand-held computer) based data entry for Industrial Surveys is outlined, using the example of a PDA-based data processing application developed recently for COSIT (Centre of Statistics and Information Technology) in Iraq. Most of the design considerations are equally applicable elsewhere. PDA data entry is done in conjunction with data entry on desktop computers and, using the COSIT example, it is shown how the two modes of data entry are used together. Java technology has been used for PDA-based application and this allows it to run on most hand-held devices (PDAs and cell phones) now available.

2.0 From Questionnaire to Tabulation

From a data processing point of view, the visible starting point for the development of any survey- based data entry programme is the survey questionnaire. The questionnaire is a key component in the 'specification document' for the data entry programme. Naturally, the more complete the questionnaire, the better the specification. However, in practice, the finalisation of the questionnaire is a long drawn-out process. Fortunately, Industrial Surveys, like other surveys, follow certain guidelines (specifically the FIRST methodology, described in the International Recommendations and originally proposed in *Strategies for Measuring Industrial Structure and Growth, UN 1994)*. Furthermore, guidelines such as these provide additional detail on data items and questionnaire contents and design. For this reason, all Industrial Survey questionnaires have some common elements which can be exploited for data processing.

In general, there are one (or more) mainly *qualitative* sections, including the general data section, and a larger number of dominantly quantitative sections (this bifurcation is useful in designing PDA-based data entry applications. The quantitative sections themselves have enough common elements in terms of validity checks and lookups for these to be generalised.

In addition, Industrial Surveys are tied to the Business Register, which is the list of all establishments in the country. One of the tasks of an Industrial Survey is to update the register and, if none exists, the census is often used to establish one. The up-to-date Business Register is subsequently used to select the list of industries surveyed. The usage and contents of the register often overlap with information on other Registers, such as the permit or licence Registers. The latter are usually maintained by separate entities and it is desirable to link the various Registers. Therefore, in designing survey data processing for enterprise data, the issue of links to the Business Register (and indirectly to the licence Register) needs to be kept in mind.

The questionnaire, together with the list of validity checks, forms the complete specification for the data entry programme. Programme-based validity checking (as opposed to manual checking done by enumerators in the field using calculators and personal judgement) can either be done in the field using PDAs or it can be done in the office. The questionnaire and the list of validity checks are specified by statistical staff and automation staff treat the questionnaire and validity checks as the requirement specification for the data entry programme.

2.1. Design Issues

Once the specification in terms of the questionnaire has been obtained, the design of the data entry system can begin. Key issues at the design stage are:

1) Number of establishments to be surveyed (determines the choice of database)

2) Choosing from technology options available

3) Integration with the Business Register, which raises the questions:

    i. Does updating the survey data automatically update the Business Register?

    ii. How does one identify the establishments which are only included in the Business Register and is uniqueness enforced?

iii. If a Business Register does not exist, is the survey conducted in two stages: stage one, for the creation of a Register, and, stage two, for establishments short-listed from the Register?

4) Coding of lookup tables such as activity and product codes (ISIC, Central Product Classification (CPC), Harmonised System), countries, units of measurement, etc. Are international standards of coding being used? If so, which version?

5) Survey procedure:

i. Is the questionnaire filled in in the field, at the office or both? If the questionnaire is surveyed in more than one stage, is there a need for an identifier to link one section of the questionnaire to the others?

ii. Central versus Regional databases: is all data entered into a central database or are there regional databases which would be merged into the central database? If there is a central database, is data entry also central or are others linked to the central database? If connected, how? And, if there are regional databases, how are they transferred to the central database?

iii. If PDAs are being used for data entry:

- How are the establishment lists allocated and distributed to the PDAs?

- How much information is entered via the PDA and how much is entered in the office?

- What is the mode of data transfer from the PDAs to the database?

6) Ownership

i. Which organization is conducting the survey?

ii. Who is responsible for organizing the data entry?

iii. Who are the users of the survey data?

iv. Is there a well-defined owner of the survey database?

Once the data entry application has been designed and implemented, like any other software, it needs to be tested and debugged before being put to final use. Testing in survey based applications is complicated by the fact that the questionnaire keeps on changing. The tested data entry application is then used in the office (or in the field when PDAs are used) to enter data. Note that data entry can be conducted in parallel with the survey or it can be

done after the survey. Questionnaires arriving at the office are not directly entered using the data entry programme. They are first coded (descriptions for activities, products, units of measurement, countries etc. are matched with their respective codes) manually and then submitted for data entry. This phase is formally referred to as the coding phase.

2.2.   Tabulation

Once survey data has been entered, it is processed and tabulated. Tabulation is the second major output of any survey data processing, the first being the data entry application. Tabulation for formal publication is normally done according to a tabulation plan, which is defined by statistical staff. In addition to the formal publication, data querying can be done on the database by any number of users. If such querying is a requirement, then a reporting or MIS (management information system) tool needs to be developed. The following sections provide practical examples of some of the salient principles mentioned in this section. and refer to actual countries and ministries in which the work was carried out.

3.0   Considerations for Survey Data Processing

Given below are important considerations that need to be kept in mind when developing Industrial Survey data processing applications in countries similar to the ones mentioned in the introduction.

3.1.   Using special purpose tools versus developing generic in-house solutions

This is one of the most important considerations in a Statistics Office when building a data processing system. Knowledge of how to use special purpose survey data processing tools (such as *Blaise* or *CS-PRO* and - although not in the same category= *SAS* and *SPSS*) can be quickly acquired by staff who do not need to have a strong background in information technology. But this approach suffers from the following drawbacks:

1) Developing data entry applications for complex surveys using these tools becomes unwieldy. The validation checking and screen-formatting options present in special-purpose tools are generalised and standardised to cover the most common uses. However, for complex surveys, the number of special cases that need to be addressed makes for inelegant solutions that are difficult to programme and maintain.

2) The proprietary database format makes data conversion cumbersome. The ease of use with which simple application development is possible with special-purpose

tools comes at the cost of a proprietary database format which does not translate easily into table-based databases. The translation is possible, of course, but it requires an additional layer of programming.

3) These tools tend to bebe superseded by new technology within a few years.

4) In some cases, such as *Blaise*, *SAS* and *SPSS*, there is an annual licence fee to be paid.

In view of these limitations and the fact that national accounts/Statistics Offices usually have dedicated ICT staff, the approach of developing generic in-house solutions is preferred in this context. The salient aspects of this approach are that it:

1) Encourages automation staff to refresh their programming skills by keeping up-to-date with new techniques and languages

2) Creates useful generic solutions to handle their data processing (more specifically data entry and tabulation) needs. Over the years, this has resulted in two practical solutions:

    i. A generic XML-based Reporting Module which can be customised to yield tabulation reports for any survey.

    ii. A template-based programming approach for data entry applications. The template is developed by the consultant using an *object-oriented* approach. This template provides the foundation for "re-use" of the template by the forms of any survey-based data entry application. This approach has been used for both normal or desktop computers and PDAs. An overview of the PDA-based data entry programme used for COSIT (Central Office for Statistics and Information Technology) in Iraq, and those of its characteristics which can be used for any Industrial Survey, is given below.

3) Works with local staff in creating and maintaining the data entry (template-based) and tabulation (generic XML based) applications. This gives local staff ownership of the applications and takes into account their limitations in designing complex data entry and tabulation applications from the ground up.

3.2. Finalization of questionnaire and testing

The most common problem in deploying data entry applications for any survey, industrial or otherwise, is the last minute changes to the questionnaire after the data entry programme has been completed. In some cases, the questionnaire is even changed while the survey is being conducted. This practice necessarily complicates the development of the data entry

programme as testing of *ad hoc* changes to the data entry programme (brought about as a consequence of the changes to the questionnaire) make the programme error-prone since there is minimal or no time available for testing.

A formal and separate testing phase for the data entry programme is necessary to ensure the quality of the data entry programme. However, because of continual changes to the questionnaire, the development of the data entry programme is also ongoing and the testing phase is necessarily delayed. Moreover, an incremental or cyclical approach, in which one phase of development-testing-feedback follows another, is not possible because testing in national accounts/Statistics Offices generally requires specialised data entry staff who are not available during the questionnaire specification and programme development stages.

There are two ways to mitigate this problem:

1) Insist on finalisation of the questionnaire before starting the design and development of the data entry programme. Unfortunately, this insistence does not guarantee that the questionnaire will, indeed, be finalised beforehand.

2) Design the data entry programme keeping in mind that last minute changes can, and will, occur. This approach requires experience in developing survey applications and programming and is not easy to teach. Nonetheless, a defensive approach such as this cannot prevent time-consuming structural changes to the programme if major revisions are made to the questionnaire once programme development has started.

3.3. Database ownership and consistency

Since there is more than one user and source of data, maintaining a unified and consistent view of the database is difficult. In the case of industrial data, sources can be surveys, licences (or permits) and industry associations. Users can be the Statistics Offices, the licensing authority, industry associations, international consultants, and the national accounts office (or sub-division). Over time, industrial data can be modified via any of the sources and there is no unified and consistent view of the data.

To overcome database ownership and consistency problems with an industrial database, an approach already tried in the Ministry of Industry (MOI) in Lebanon may be useful and is as follows:

1) Designate a database owner (n the case of Lebanon, it is the Ministry of Industry)

2) The database owner maintains a consistent and up-to- date industrial database. A unique ID scheme is enforced (in the case of Lebanon, the unique ID is composed of

a 4-digit ISIC, 7 digit region-code and a serial number within the ISIC-Region combination).

3) Create interfaces for the different sources to enable them to update specific sections of the industrial database (for Lebanon, this has still to be done but the proposal consists of one interface for Permits and possibly another one for the Associations of Lebanese Industries or ALI).

Over time, it is probable that the users, too, will want to update some features of the unified industrial database. So long as there is a single owner, any number of update interfaces using the Internet (or a wide area network) can be provided by the owner, whose responsibility it is to compile a consistent and up-to-date industrial database.

3.4. Technology options

The technology choices outlined in this section assume the development of in-house solutions and that no special-purpose tools are used. 3.4.1. Database

For survey applications, anything from lightweight database applications, such as MS-Access and Paradox (currently owned by Corel; previously owned by Borland), to server-based systems such as SQL Server, Oracle and MySQL can be used. The choice depends on a number of factors, described in detail below.

Expected number of records
Practical experience suggests that a good way of determining the size of the database in choosing between a small one (MS-Access, Paradox) or a large one (MySQL, Oracle, SQL Server) is this: if the expected number of records in a survey is around a few thousand, then a small database is sufficient; otherwise, a large server-based database is more suitable. Generally, a survey application database is distributed over a number of database tables so the rule could more correctly be stated to apply to the expected number of records in the *largest* table. That said, Paradox tables have been used for much larger databases with several million records but, with such a high number of records, file corruption problems may start appearing.

Number of concurrent users for data entry
A high number of concurrent users (greater than three as a general rule although small databases can support a much higher number in principle) is a good indicator of the need to use a server database since they are built for heavy use.

Pre-existing licences
If a National Statistics Office has existing licences for a particular database, then there is an automatic bias towards using that database for survey processing regardless of other

concerns. If no licensing agreements are in place, Paradox is a good choice for small databases (mainly for its querying facilities) and MySQL is good for larger ones (as it is a robust and popular server database available free of charge.)

User querying/reporting requirements

Frequently, statistical and national accounts staff (among others) need to query the database independently (i.e., without having to go through the automation staff). Rather than preparing specifications for the automation staff every time the subject area users need to get some information from the database, it is useful if the users themselves have the capability to extract information. There are two practical ways that users can query the database on their own:

1) Train users on Paradox QBE queries

Where small database usage is called for, the Paradox application has a very practical visual query tool called QBE (Query by Example) that can enable users with some training to create powerful queries. The querying capability acquired with some training can satisfy most users' needs. Note that SQL (Structured Query Language) is the most common language used to extract and manipulate data in all the databases available today. But learning SQL for anything beyond the simplest tasks is quite difficult and, generally, such training is best suited to technical staff.

2) Using customised query tools. Custom query programmes for users can take the form of:

    i. Programmes that run customised SQL or QBE scripts and prompt the user for some parameters

    ii. A general purpose reporting tool

In-house database automation expertise

Availability of in-house automation expertise in a particular database does, of course, create a bias in the choice of database to be used for a survey. More significant is the decision which has to be taken when very little or no expertise is present. In such cases, using MS-Access, which is familiar to almost everyone and is present in most offices, is a practical choice (unless of course, other factors dictate using a server database, in which case some database administration training needs to be given).

Choice of development technology

Certain databases are natural fits for particular development tools. For example, Microsoft .NET technology is easier to use with Microsoft databases. Similarly, Paradox is very well suited to the Delphi programming environment. The importance of this criterion is determined by the availability of application programme interfaces and their support by the

appropriate drivers (ODBC – Open Database Connectivity and JDBC – Java Database Connectivity).

Modern programme environments allow convenient access to many different database management systems. For object-oriented programming languages like C++ and C#, Java class libraries are available. These provide methods for access to many different databases, thus alleviating the programmer from having to consider the database specific details.

Although ODBC is available on many operating systems, it is used primarily on Microsoft Windows. Currently Microsoft ships ITSW (Information Technology and Software) own ODBC drivers with any supported Windows version. In the examples given in Section 4.0 below, the database connectivity is achieved using ODBC.

Application development (programming)

The choices for application development (programming) are many. Listed below are examples of those that have been used over the years, the factors influencing the choices, and the considerations for using them in surveys.

Borland Delphi Win32 applications (Pascal language)

Delphi has been one of the foremost application development environments over the years. This tool works particularly well with the Paradox database. Thus, for surveys with a relatively small number of records and concurrent users, Delphi has proved to be a very productive tool. The original Delphi programming was done using the older Win32 model of the Windows operating system. For applications using the Arabic language (used in Lebanon and Iraq), limitations in the Win32 programming model were encountered and these led to the choice of another more recent component of Delphi, namely the Dot Net (.NET) application development component (which uses the Windows Dot Net programming model) which is discussed next.

.Net applications (C# language, Visual Studio)

.NET is a software platform developed by Microsoft. It consists of a run time environment, class libraries for the programmer (APIs - application programme interfaces) and services. This platform in its entirety is only available under MS Windows (versions for different Unix systems are available and supported by the projects Mono and *dotGNU*). The platform provides flexible methods for accessing operating system level services and replaces old concepts like COM. It is possible to use .Net for development for device platforms like smart phones and PDAs.

The Common Language Infrastructure (CLI) of .Net supports many languages which are already available in the Microsoft platform. Almost all these languages compile with a Common Intermediate Language, an intermediate language, which can be executed using an implementation of Common Language Infrastructure (CLI) delivered as part of the MS .Net framework.

.Net provides a modern implementation of the new industry standard Web services – there is an infrastructure for distributed programming in LAN as well as in Internet.

The most widely used CLI language is C#, which bears similarities to C++ and Java and is intended to be a simple, modern, general purpose, object-oriented programming language.

Visual Studio is the integrated development environment offered by Microsoft for development with the CLI languages under the .Net framework. The current version 9.0 (Visual Studio 2008) supports Visual Basic, C++/CLI and C#. Visual Studio is also available in the *Express* edition (a free of charge, limited version of Visual Studio 2005 and 2008) which supports only one programming language. This Express edition is available only for a limited time period of 30 days but, if the user is registered with Microsoft, the time period is extended.

Java

Java technology consists of a platform-neutral programming language and a set of tools. UNIDO has used Java over the years for many tasks (surveys and other uses) and for Industrial Surveys in particular Java has been used for the following purposes:

1) *Data transfer from special format to database*. Java has very good 'low-level' data manipulation capabilities (which Delphi lacks) and these capabilities make it a very useful tool for transferring data from one format to another. For example, data saved by the PDA data entry application need to be transferred to an SQL Server (this was done, for example, in Iraq,) and this is achieved via a Java application. Similarly, the lists of units that each PDA is designated to cover are generated by a Java programme.

2) *PDA data entry programme*. Java comes in three editions: Standard, Enterprise and Micro. The Micro edition of Java is used to develop applications for hand-held devices (PDAs or Portable Digital Assistants). Since Java is platform neutral, it is possible to use a programme developed in Java on any number of commercially available PDAs or mobile phones. For Industrial Surveys, a PDA-based data entry application was recently developed in Java for use in Iraq*;* specifically with the Java MIDP 2.0/CLDC 1.1 specification. The platform neutrality of Java has made it especially useful since   any of a number of available portable devices and mobile phones can be selected for use, depending upon their screen size, memory and cost.

Perl

Perl is a programming language which is very effective in writing quick applications for string formatting required in data format conversions in survey applications. Format conversion is a common requirement in all data related applications as data from disparate sources need to be put under a single umbrella.

4.0. Generic Survey Reporting Module

As part of UNIDO sponsored work in Saudi Arabia with the Ministry of Commerce and Industry in 2005, a survey reporting module was designed and developed that would be flexible enough to provide tabulations from databases compiled from surveys. The idea is that, for any survey, the technical staff would create the specification file (described below) and users could use the programme to query the survey database without needing any special training (see the User Manual in Annex 1).

Three types of tabulation are possible:

1) Simple classification: this type of classification requires one classification parameter (such as ISIC-3 digit, city, company type, etc.) and one or more value parameters (such as number of employees, taxes, wages, etc.).

2) Cross classification: this tabulation requires one row parameter, one classification parameter, and one value parameter.

3) General data: this report allows one or more general data parameters to be printed out (such as name, address, phone number, etc.).

Reports are generated in HTML format and are displayed on screen and can be printed out and saved. Saved reports can be imported in a spreadsheet. The generic and flexible nature of the system means that the system can produce tabulations from any database that follows some assumptions (as described in Annex 1). Each database is described (in terms of its classification, value and general data parameters, their field and table mappings, and lookup tables) using a specification file, which is in the Extensible Mark-up Language (XML) format. Given below are the elements of the XML specification file (see Annex 2 for a sample XML specification file).

4.1. The Specification File

The survey reporting system works off a specification file that describes the following:

- The database and its ODBC alias

- The main table and its primary key

- Classification parameters with optional lookup tables

- Value parameters

- General data parameters

- Filters

These specifications need to be written out in XML format. XML allows specification of files in a hierarchical fashion (that is, elements can be embedded within other elements). There can be only one primary or document (root) element under which all other elements are defined. XML is case-sensitive, so the user should follow the alphabetic-case for all elements and attribute names specified below.

Note that all element and attribute names are specified in italics.

4.2.   Reporting system

The document (root) element is named *ReportingSystem* and has the following attributes:

- *Survey*: display-only attribute used to name the survey which is displayed on the reporting system screen

- *Year*: display-only attribute used to identify the year of the survey

The *ReportingSystem* element has the following nested elements (or children) which specify the rest of the system.

4.3.   Main DB

The *MainDB* element has the following attributes:

- *Name*: the name of the database which is a display-only attribute

- *MainTable*: the name of the main table that is supposed to have:

    i.   All the classification parameters, and

    ii.   All the general data parameters

Important note: The assumption that all the classification and general data parameters are derived from a single table, namely the one specified by the *MainTable* attribute, is significant. Another assumption is that each row of the main table is unique. Sometimes, database design is done in such a way that classification parameters are distributed over more than one table, all having the same primary key. In such a case, the database administrator should be asked to create a 'view' or temporary table that has all the

necessary classification fields in order for the reporting system to access all classification parameters.

- *ODBC*: this is the name of the ODBC alias for the database

In addition to the above attributes, the *MainDB* element has one or more nested *Key* elements. Each element specifies a field of the primary key for the main table. Note that the sequence in which the key fields are specified is important as it should match the sequence of key fields in the main table. Each *Key* field has a *Name* attribute which is the name of the key field in the main table.

4.4.   Parameters

The Parameters element has three sub-elements corresponding to the three types of parameters used in the Survey Reporting System: classification, value, and general data.

4.4.1.1.   Classification

The *Classification* element has one or more *Parameter* elements, each with the following attributes:

- *Name*: the name, which is displayed in the programme and used as the title for the column (or row) in the reports

- *Field:* the name of the field in the *MainTable*

In addition, the *Parameter* element corresponding to the *Classification* element can have an additional *Lookup* element with the following attributes:

- *Table:* the name of the lookup table

- *KeyField:* the name of the field in *Table* above corresponding to *Field* in *MainTable*

- *LookupField:* the name of the lookup field, which actually shows up as the title for the column (or row) in the reports.

Important note: All the Parameter elements in the Classification section are fields which are assumed to belong to the *MainTable*.

4.4.2   Value

The *Value* element again has one or more *Parameter* or *Group* elements. The *Parameter* element corresponds to an individual value parameter (SQL note: it is going to be 'SUM'med

in the 'SELECT .. GROUP BY' query) in a table, whereas the *Group* element corresponds to a table with a number of individual *Parameter* sub-elements. Each *Parameter* element under the *Value* element has the following attributes:

- *Name:* the name of the value parameter appearing in the programme and used as the title for column (or row) in the reports

- *Table:* the name of the table in which the value parameter exists

- *Field:* the name of the field in *Table* corresponding to the parameter

In addition to the above attributes, the *Parameter* element within the *Value* element has one or more nested *Key* elements. Each element specifies a part of the foreign key corresponding to the primary key in the main table.

The *Group* element under the *Value* element is similar to the *Parameter* element but with the following differences: i) the common (to all nested parameters) *Table* attribute and (nested) *Key* sub-elements belong to the *Group* element rather than its nested *Parameter* elements; and ii) the nested *Parameter* elements only have two attributes: *Name* and *Field* (since the *Table* attribute and *Key* sub-element are defined in the parent *Group* element).

Important note: A special *Parameter* under the *Value* element is used for which the *Table* attribute has the value "N/A" and the *Field* attribute has the value "Count". This special value parameter is used to count the number of units in the main table (e.g., the number of establishments). In general, a *Value* parameter with "N/A" for the value of the *Table* attribute indicates that the parameter is in the main table.

4.4.3 General data

The *general data* element has one or more Parameter sub-elements (corresponding to the general data parameters), each with the following attributes:

- *Name*: the name which is displayed in the programme and used as the title for the column in the reports

- *Field:* the name of the field in the *MainTable*

Important note: All general data parameters are assumed to belong to the main table.

4.5.   Filters

The *Filters* element has one or more *Filter* elements with the following attributes:

- *Name:* the name of the filter appearing in the programme and report

- *Condition:* the SQL condition that is appended to the WHERE clause in the generated SQL statement to filter the chosen records

5.0   Using PDAs for Industrial Surveys

Traditionally, survey data are entered in questionnaires, which are then sent to one or more offices to be entered into the database. The point at which data validation takes place is in the office(s) where the data is entered into the database. In many cases, at this stage of data validation, feedback is required from the source of data (factories and head offices in the case of Industrial Surveys). Getting feedback is normally both difficult and expensive, or not possible at all. Hence, it is desirable to enter data in the computer in the field so that data validation can be done on site. PDAs or hand-held computers are a relatively inexpensive way to accomplish this task.

Recent developments in mobile phone technology have made it possible for mobile phones to serve as PDAs for data entry. The presentation later in this chapter is based on the experience gained from two projects: i) a consumer price index data entry application which has been in use for more than three years; and ii) a data entry application for an agro-industrial survey for COSIT (Centre for Statistics and Information Technology) in Iraq.

5.1   Data entry process: questionnaire, PDA and desktop

1) Before the fieldwork begins, individual PDAs are pre-loaded with lists of those establishments which they will cover.

2) In the field, data are always first entered in the paper questionnaire.

3) Once the questionnaire is completed, numerical data (as opposed to general data and other qualitative data) are entered via a PDA and validated.

4) Data entered on the PDAs are transferred to the central (or regional) databases via files which are stored on the PDA (this is the approach followed in Iraq) or they are transferred on-line using the Internet (this presupposes PDAs with an Internet connection).

5) Once PDA data have been transferred to the database(s), data for the non-numerical sections and those sections which require coding (product codes, activity codes, country codes, etc.) are entered manually via a desktop data entry programme.

6) [Optional] If more than one database (for example, one database per region) exists, they need to be merged into a central database.

The reasons why data are always entered first in the paper questionnaires are:

- To maintain a paper trail

- To enable coding at the office

- To enter non-numerical data (using the desktop system) which is not covered by the PDA

Having a complete desktop based system is necessary, in addition to having the PDA data entry programme, for the following reasons:

- Entering textual data on PDAs is cumbersome because of the limited screen size and the small keyboard

- Coded sections can only be entered once the questionnaires have been coded at the office(s)

- Updates of questionnaire data are best done from a central database using the desktop programme

A short description and a flowchart of the data entry programme are given below and illustrated by some screenshots of the application. The options below are applicable to any industrial or other type of economic survey.

Figure 1: Survey data entry scheme

Survey data entry scheme

Numeric values — On PDA in the field

Editing of numeric values — On PDA in the field

Coding of classifications, etc. — Manually, in the office

Numeric values — On Desktop in office

Data entry of non-numeric values — On Desktop in office

Editing of information — On Desktop in office

Clean data set for tabulation and analysis
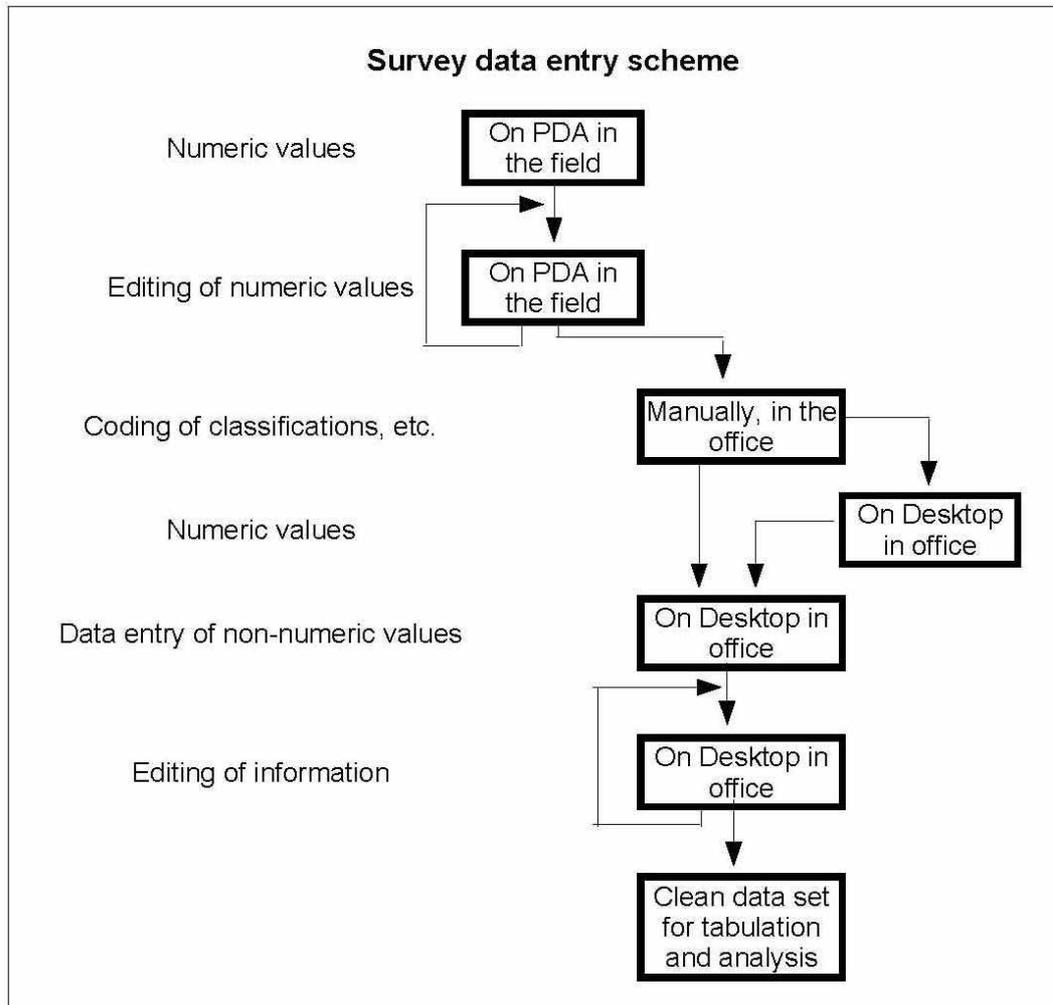
Figure 2: Start-up Screen

Figure 3: Search/List Menu

5.2.   Data Entry Programme: Main Menu Options

5.2.1   New Form

he *New Form* operation creates a new questionnaire with a chosen ID (for Iraq, the ID is a combination of the ISIC and serial number). First, the user is asked to look up or search for an ID. After an ID is chosen, the data entry screen for the first numerical section (which is the first section of the questionnaire to be entered on the PDA) appears. The user can then proceed with data entry as described below in 5.2.7 on *Data Entry, Validation, and Saving*.

Before the data entry screen for the first section appears, the programme checks whether the form for the chosen ID has been created before. If this is the case, then the user cannot create another form with the same ID and he/she is told so (to change an existing form, the user needs to choose the menu options *Retrieve Form* and *Edit Current Form* described next).

In addition, before allowing the user to create a new form, the programme checks whether an existing form is already being entered. If so, the programme ensures that all the sections of the questionnaire have been saved and does not allow the user to create a new form. Note that, as each screen is exited using any menu option other than *Cancel*, the form is saved if there are no errors on the screen (see 5.2.7 *Data Entry, Validation, and Saving* below).

Once the user has finished entering data in this section, he/she can move on to the next screen by pressing the button on the menu called *Next.* It is also possible to jump to any section of the questionnaire by choosing the button on the menu called *Go To.* Navigation among sections is described in 5.2.8 *Navigating Screens* below.

5.2.2   Edit Current Form

This operation is similar to the *New Form* operation since both take the user to the first data entry screen. However, the difference is that for *Edit Current Form*, the programme goes directly to the questionnaire which has either been recently created (using *New Form*) or retrieved (using *Retrieve Form.*) This is useful when the entire questionnaire is not entered in a single session and the data entry operator resumes entering data on the same questionnaire at some later point in time. When using *Edit Current Form*, the user does not have to first use *Retrieve Form* in order to call up the current questionnaire.

The same data entry, validation, saving and navigation rules apply for this operation as for *New Form*.

5.2.3.   Retrieve Form

Questionnaires which have already been entered can be retrieved for modification at any point in the future by using this menu operation. On choosing *Retrieve Form,* the user is shown a list of questionnaires which have already been saved. He/she can then choose one

from this list and that questionnaire becomes the 'current' form, ready for modification using the operation *Edit Current Form* as described above.

Note that, if there are one or more sections of an existing questionnaire which have not been saved, the user is not allowed to retrieve forms and, in this case, must use the *Edit Current Form* function to access that questionnaire.m

5.2.4. Import ID List

Before the user can do any data entry, the list of questionnaires which can be entered on the PDA needs to be imported. This list is saved in the *ID.dat* file which should have been transferred to the predefined location specific to the device.

The *Import ID List* operation need only be carried out when a new list of questionnaire IDs is made available (via a new *ID.dat* file.)

5.2.5. Export Forms

The data which has been saved on the PDA eventually needs to be transferred to the computer. The *Export Forms* operation is the first step in the transfer process. Upon choosing this operation, all the existing forms are transferred to a data file which lies in the same directory as the *ID.dat* file. Note that the existing data forms on the PDA remain intact. If there is an existing data file on the phone, that file is overwritten with the new data.

5.2.6. Delete Current Form

Using the *Delete Current Form* operation, the user can delete the form currently being entered regardless of whether it is new or retrieved.

5.2.7. Data entry, validation and saving

The data entered on all screens during data entry are verified for accuracy before saving. A number of checks have been put in place, most of which have to do with summing together a number of fields and checking whether the sum is the same as that of a total field on the same form. If the validity check is not met, the user cannot exit the screen except by cancelling (using the *Cancel* menu button). All other menu options first validate and then save the contents of the screen to the PDA memory. Only when the user chooses *Cancel* does the programme bypass the validity check and saving procedure.

5.2.8. Navigating Screens

From any data screen, the user has these two basic buttons on the menu:

- *Done*

- *Cancel*

Done goes back to the main menu after validating and saving the form, whereas *Cancel* takes the user back to the main menu without any validation or saving. In addition to the above two basic buttons, the following navigation options are available:

- *Next*

- *Prev*

- *Go To*

- Section specific screen buttons (corresponding to table rows)

The *Next* and *Prev* buttons take the user to the next and previous screen in sequence (relative to the current screen) respectively. The *Go To* button allows the user to jump to a number of locations in the questionnaire. These locations correspond to sections or sub-sections of the questionnaire. In addition to these three buttons, some screens also allow the user to go to screens within a section, especially those sections where each screen corresponds to a row in the table.

Given below are the scenarios in which both the PDA data entry programme and the desktop data entry programme work in conjunction with one another.

5.3. Data entry scenarios

The desktop data entry programme can be used for data entry exclusively or in conjunction with a hand-held device.

5.3.1. Exclusive usage of a desktop programme for entering establishment data

This refers to using the desktop data entry programme for those industrial units whose data has not been entered using the PDA. In this case, the steps involved are as follows:

1) Entering new general unit data in the Register [optional - if unit data have not been uploaded previously in the Register]

2) Updating bio-data information

3) Transferring unit from Register to Survey database and assigning a questionnaire ID (note this assumes that a central Business Register exists from which the survey units are drawn; this assumption is made in the case of both Lebanon and Iraq.)

4) Entering the numerical survey questionnaire data.

Note that this scenario assumes that data entry forms have been coded beforehand.

5.3.2. Entering data using desktop and hand-held devices

This refers to using the desktop data entry programme for those industrial units where the numerical data have already been captured by the PDA and transferred to the database. In this case, the steps involved are:

1) Updating general data information

2) Transferring unit data from the Register to the Survey database and assigning a questionnaire ID

3) Updating code, unit and quantity data

4) Entering control information (enumerator, supervisor etc.)

5) [Optional] Entering additional qualitative data

This scenario assumes the following:

- Register data have been populated beforehand

- Numerical data for selected units have been captured by the PDA and transferred to the database

- Data entry forms have been coded

5.3.3. Updating data using desktop data entry programme

This refers to using the desktop data entry programme once all the data have been entered using either of the two options above. The steps involved are:

1) Navigating to any part of the questionnaire with the user interface

2) Updating data for the section

Annex 1:   Survey Reporting System: User Manual

A1.1   Introduction

The reporting module enables a user to generate tabulations of interest from any industrial database. The system is database-independent, meaning that any industrial database (defined in next section) can be used for generating reports. Each new survey is defined using a specification file in XML format.

The types of reports that this system produces are as follows:

- *Simple Classification*: User chooses one classification parameter, and one or more value parameters (limit of value parameters is ten; classification and value parameters are further defined below).

- *Cross Classification*: User chooses one row classification parameter, one column classification parameter and one value parameter.

- *General data*: User chooses one or more general data parameters (limit is five; definition of general data parameter is given below).

A filter to restrict the records to be processed can be applied to any of these reports.

A1.2   Output format

The reports are generated in HTML format which may be:

- Viewed on screen

- Printed

- Exported to a spreadsheet

A1.3   Assumptions and definitions

In order to keep the reporting system general enough to accommodate any 'industrial' database, it is necessary to restrict the functionality of the application to generate reports

which are of interest from a national accounts perspective for an industrial database. The following assumptions are held by the programme:

- *Single database*: All information is derived from a single database.

- *Main table and primary key*: There is a 'main table' which has a primary key composed of one or more fields.

- *Classification parameters*: Classification parameters are individual fields from the main table. Optionally, each classification parameter may have a corresponding lookup table.

- *Value Parameters*: Value parameters are numerical fields. They are in general not from the main table (i.e., from tables which are linked to the main table with a -foreign- key corresponding to the primary key). However, they can be from the main table. A special value parameter, 'Count', is used to count the number of units in the main table grouped according to the chosen classification parameter (or row and column parameters for cross-classification reports).

- *General data parameters*: General data parameters are individual fields from the main table. Optionally, each general data parameter may have a corresponding lookup table.

- All reports can be subject to a filter, which is a set of conditions imposed on the records in the main table.

- The database, its tables and fields (classification, value and bio-data), and the filters are all specified in an XML file called the specification file, and whose format is given in Annex 2 of this chapter.

A1.4 Required files, installation, and un-installation

Required Files

The reporting system application can be installed in any directory (chosen typically by the installation programme). For the sake of generality, we assume that the directory is referred to as <install_dir>.

The following are the files required for running the survey reporting system:

- <install_dir>\MCISurveyReport.exe (the executable file)

- <install_dir>\setup.ini

- <install_dir>\output\ (directory for storing output files)

Installing from a CD Insert the installation CD in the drive

Carry out the following two steps:
1) Locate the programme setup.exe

2) Run the programme by double-clicking on it, or selecting it and pressing the *Enter* key

You will then be asked to follow a series of steps, including one in which the installation programme prompts for the installation directory (<install_dir>), which you may change from the supplied default if required. All required files mentioned above are copied to the installation directory. Note that the installation CD (or USB drive) will also have one or two XML specification files copied to <install_dir>. In addition, required databases defined in the specification files and their ODBC aliases are all created automatically by the installation programme.

Finally, a short cut to the survey reporting programme is created on the desktop. The exact name will be survey, statistics office, or country specific, and is determined by the installation programme.

setup.ini file (for direct copy only)

The specification file is named in the setup.ini file and, with the possible exception of first-time use (if the required files are directly copied and not installed), it is not used by the user but changed (if necessary) only by the programme.

If you have copied the required files from a CD or USB drive instead of installing, then (and only then) before using the reporting module for the first time , you have to change the value of <full path for XML specification file> to point to a valid specification file.

```
[Specification]
SPECFILE=<full path for XML specification file>
```

Specification files

The reporting system is intended to work with any Industrial Survey database; it is kept general by keeping all survey specific details in a specification file, which includes the following:

- Name of survey

- Year of survey

- The database name and its ODBC alias

- Tables, keys and lookup information for classification, value and general data parameters

Note that the name of the specification file in use at any point in time is given in the setup.ini file above. The ODBC alias for the database should be set up before running the programme. The installation programme does this automatically but, if you copy the required files manually, then instructions for setting ODBC aliases are given below.

Details of the specification file, which is in XML file format, are given in Annex 2: Sample XML Specification File for Reporting Module. The specification file is typically prepared by a technical person (normally a programmer or systems analyst who has read and understood the specification file format) along with a subject area specialist (a statistician, economist or national accountant).

ODBC Aliases (for direct copy only)

Each database to be tabulated by the reporting system needs to have an ODBC alias that is defined in the specification file corresponding to that database. Before running the reporting system, the necessary alias (or aliases if more than one survey is to be tabulated) must be in place. The installation programme automatically sets up these aliases for you. If you are copying the required files directly, you would be required to set up the ODBC alias (or aliases) manually. The steps required for setting up an ODBC are as follows:

1) Choose [Control Panel]/[Administrative Tools]/[Data Sources (ODBC)] from the Start menu

2) Press the [Add] Button

3) From the displayed list of available 'drivers', select the appropriate driver and press the [Finish] button ('Microsoft Access Driver (*.mdb)' would be the most usual but it could be any of the available ones).

4) A screen appears showing the database specific parameters which need to be filled up. For Microsoft Access databases, you need to enter the [Data Source Name], making sure that it is identical to the name entered in the specification file, and then press the [Select...] button to browse and locate the Microsoft Access database file and press [Ok].

Uninstalling

Carry out the following three steps:

1) If installation was done by an installation CD or USB driveGo to [Settings]/[Control Panel]/[Add Remove Programmes]

2) Select the survey reporting programme from the list (the exact name will be the survey statistics office or country specific)

3) Press the [Remove] button

If installation was done by copying required files and manually setting up required options, then the following steps are necessary:

1) Empty the contents of <install_dir>

2) Remove the databases, if they were not already removed, by emptying <install_dir>

3) Remove the specification files, if they were not already removed, by emptying <install_dir>

4) Remove all ODBC aliases that were created for the reporting programme

A1.5 Generating reports: overview

The steps for generating reports s as follows:

1) Choose parameters

- *Simple Classification*: User chooses one classification parameter and one or more (up to ten) value parameters.

- *Cross Classification*: User chooses one row parameter, one column parameter, and one value parameter.

- *General data parameter*: User chooses one or more (up to five) general data parameters. The order in which they are chosen determines the sort order in the resulting report.

2) Choose filter

3) Generate report

A1.6 Starting up

Figure 4: Start-up screen

Upon clicking the MCISurveyReport.exe file in the <install_dir> (or on the short cut), you should see the following screen.
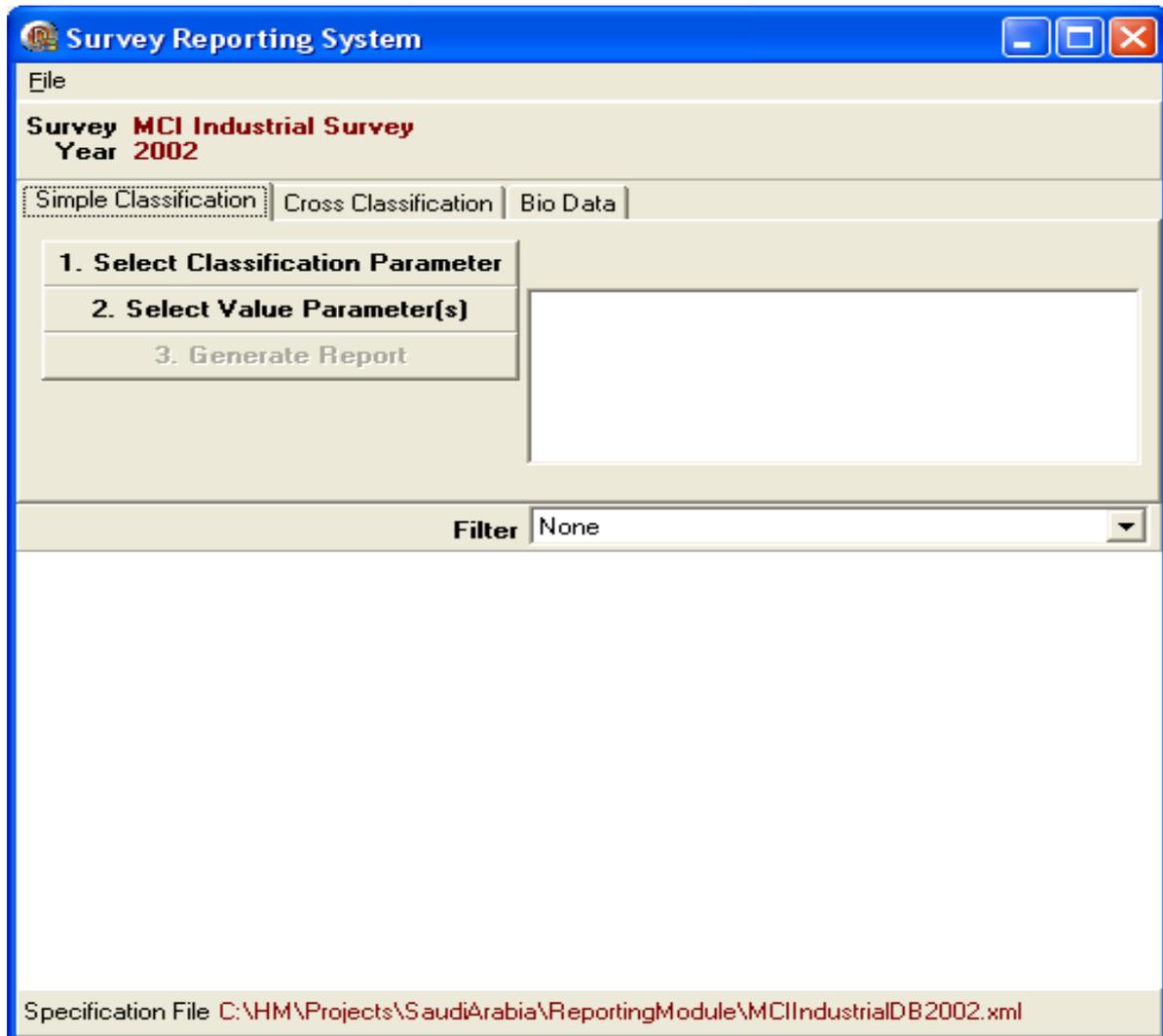
Note that the survey name and year are given on the top part of the screen and the name of the specification file is given at the bottom. The empty part of the screen at the bottom is where the report is displayed in HTML format. Right on top of the display area is a pull-down menu labelled 'Filter'. Above the filter is a section that has three tabs (or sections) corresponding to the three types of reports that can be produced by the Reporting System: simple classification, cross classification, and general data. Generating these reports is described in more detail below.

A1.7 Simple classification report

As Figure 5 indicates, there are three options for generating a simple classification report:

1) Choose a classification parameter by pressing the [*1. Select Classification Parameter*] button

2) Select one or more (up to ten) value parameters by pressing the [*2. Select Value Parameter(s)*] button

3) Optional: Choose a pre-defined filter from the [*Filter*] pull-down menu.

4) View the report by pressing the [*3. Generate Report*] button.

Upon pressing the [*1. Select Classification Parameter*] button, the following dialogue box shows up:

Figure 5: Classification Parameters
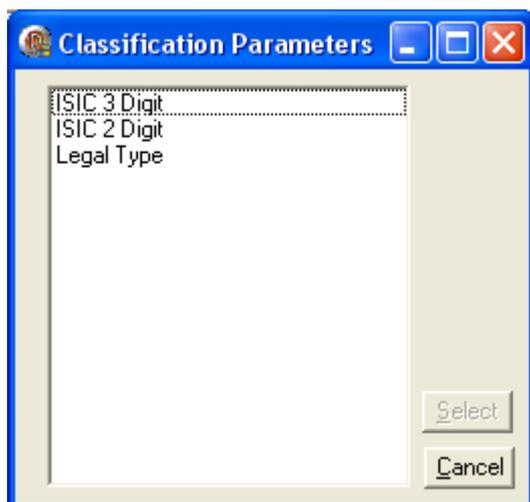
The list box in this dialogue box lists the classification parameters than can be chosen for this survey. Note that these classification parameters are survey specific and are given in the specification file.

Only one parameter from the list can be chosen and, upon doing so, the [*Select*] button, which is disabled in Figure 2, becomes enabled. Pressing [*Select*] takes the user back to the main screen and the chosen parameter is displayed next to the [*1. Select Classification Parameter*] button.

At this point, the user needs to press the [**2. Select Value Parameter(s)**] button and, upon doing so, he/she will see the following dialogue box:

Figure 6: Value Parameters
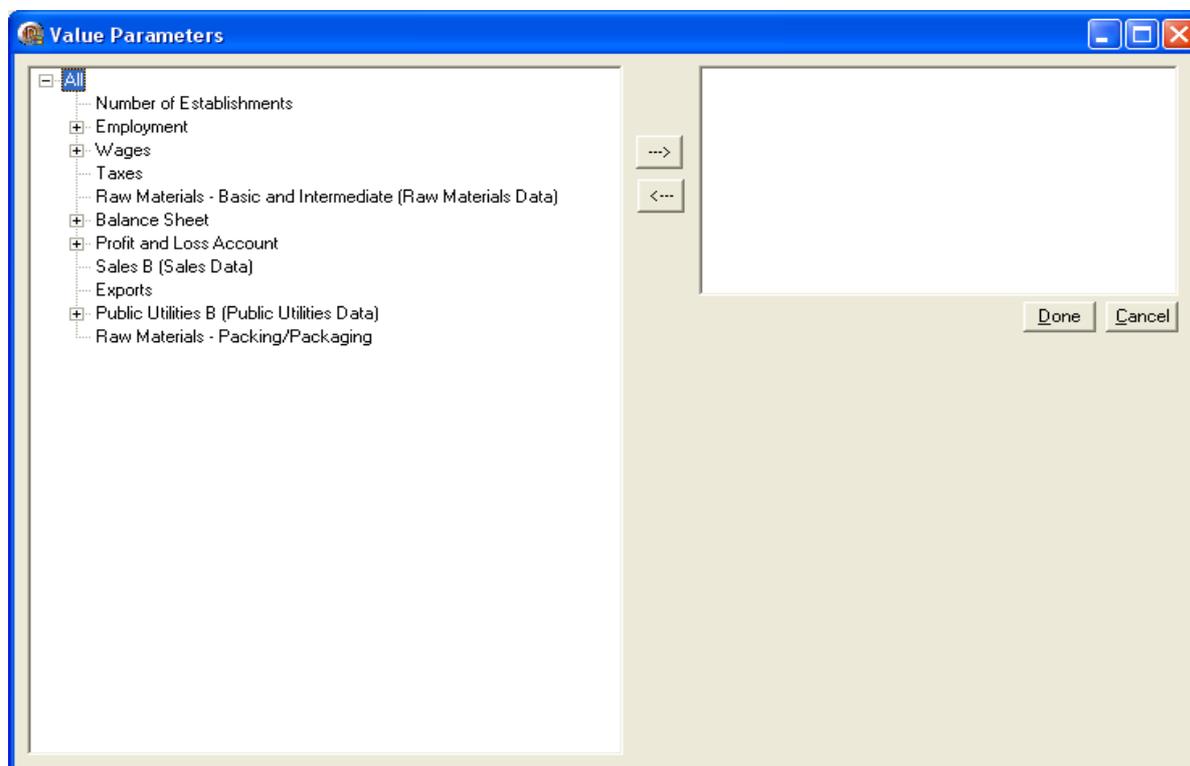
The box on the left hand side lists all the value parameters that are applicable to this survey as specified in the XML specification file. The box on the right hand side shows the list of chosen value parameters. The parameters with the [+] sign on the left are grouping parameters (more strictly speaking, they belong to the same table in the database) and they can be expanded. Selecting one or more of the value parameters and pressing the [->] button will copy the parameters to the right hand box of chosen value parameters. Likewise, pressing the [<-] button after selecting a chosen value parameter will remove it from the chosen list.

Note that you cannot choose a grouping parameter (those marked with a [+] sign on the left) to be included in the chosen list on the right. Clicking on [+] next to a grouping parameter would expand the list and the [+] sign changes to [-] indicating that it can be collapsed by clicking on it again.

The figure below shows the same dialogue box as above but with a few chosen parameters and the 'Profit and Loss' section expanded.

Figure 7: Value Parameters – Expanded tree
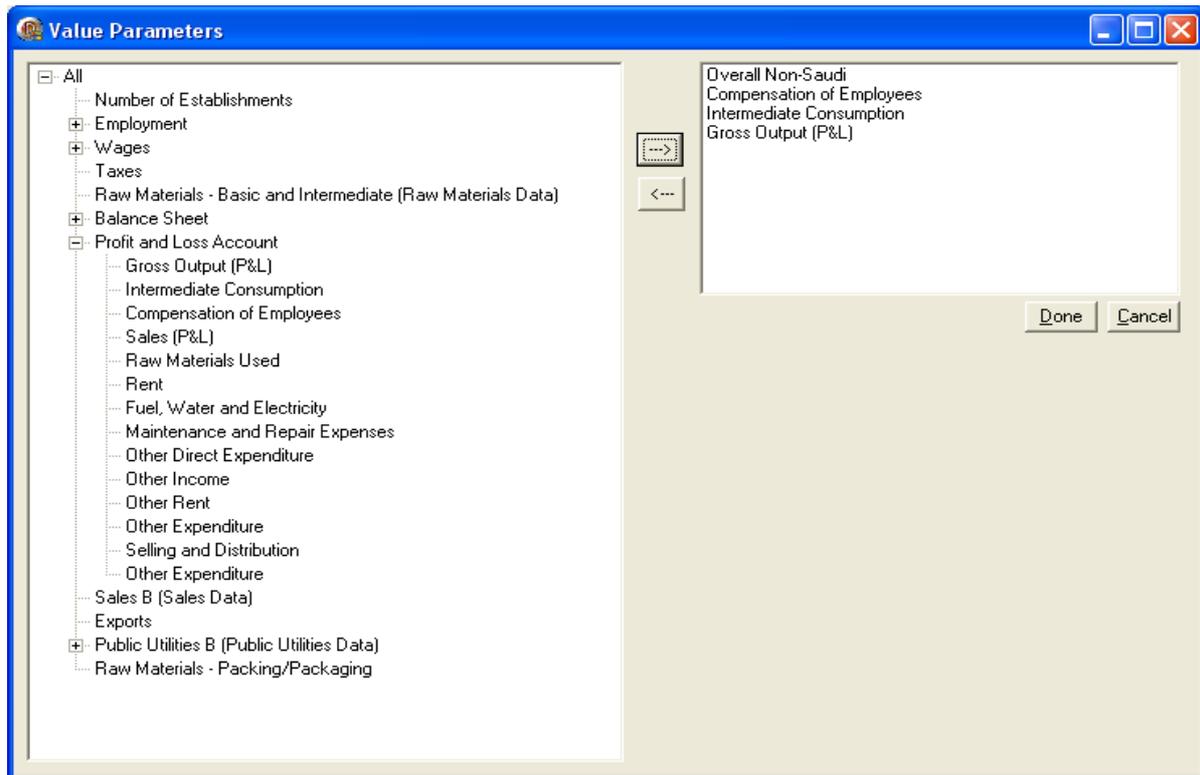
Once the classification and value parameters have been selected, the report can be viewed and printed by pressing the [*3. Generate Report*] button. Figure 8 below shows an example of the resulting screen, which has been maximized:

Figure 8: Simple Classification – Generated Report
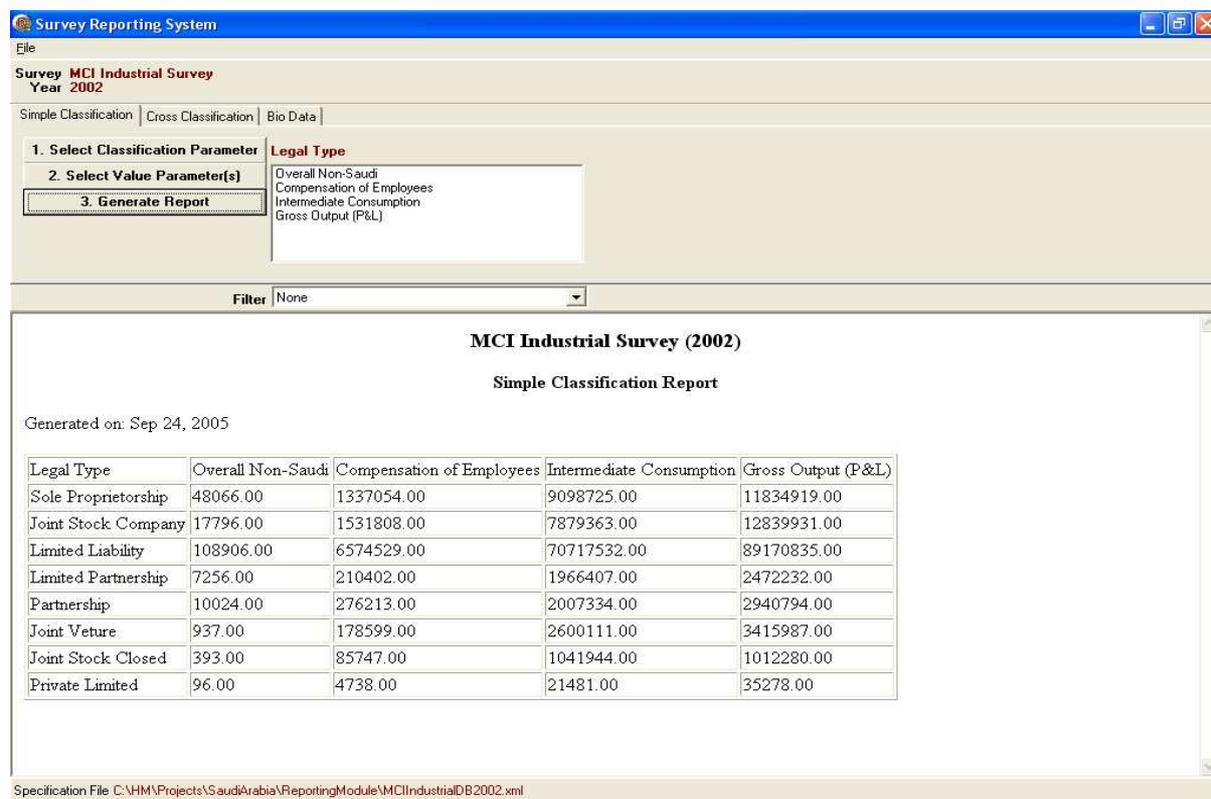
The report can be printed by right-clicking on the report display window and choosing [**Print**].

A1.8 Cross classification report

The procedure for generating the cross-classification report is similar to the simple classification report. After choosing the [**Cross Classification**] tab ([**Simple Classification**] is the default tab), a series of simple steps need to be followed:

1) Choose a row classification parameter by pressing the [**1. Select Row Classification**] button

2) Choose a column classification parameter by pressing the [**2. Select Column Classification**] button

3) Select one (and only one) value parameter by pressing the [**3. Select Value Parameter**] button

4) Optional: Choose a pre-defined filter from the [**Filter**] pull-down menu

5) View the report by pressing the [**4. Generate Report**] button

The dialogue boxes for choosing the classification parameters are as described in the section for simple classification.

Below, a cross-classification report is shown with 'ISIC 2 Digit' as the row classification parameter, 'Legal Type' as the column parameter, and 'Intermediate Consumption' as the value parameter.

Figure 9: Cross-Classification Report
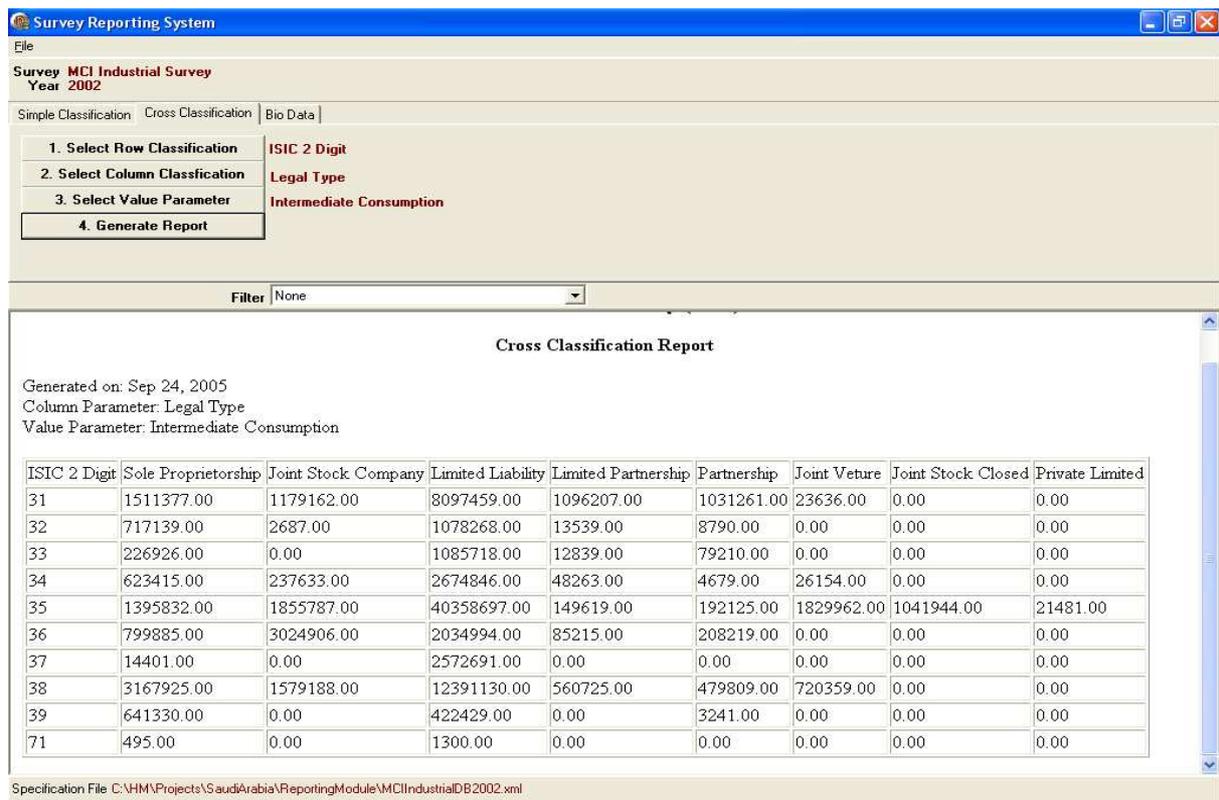
A1.9 General data report

Generating general data reports requires the following steps after choosing the [**Bio Data**] tab:

1) Choose one or more (up to five) bio-data parameters by pressing the [**1. Select Bio Data Parameters**] button

2) Optional: Choose a pre-defined filter from the [**Filter**] pull-down menu

3) View the report by pressing the [**2. Generate Report**] button.

Pressing the [*1. Select Bio Data Parameters*] button shows up the following dialogue box:

Figure 10: General Data Parameters
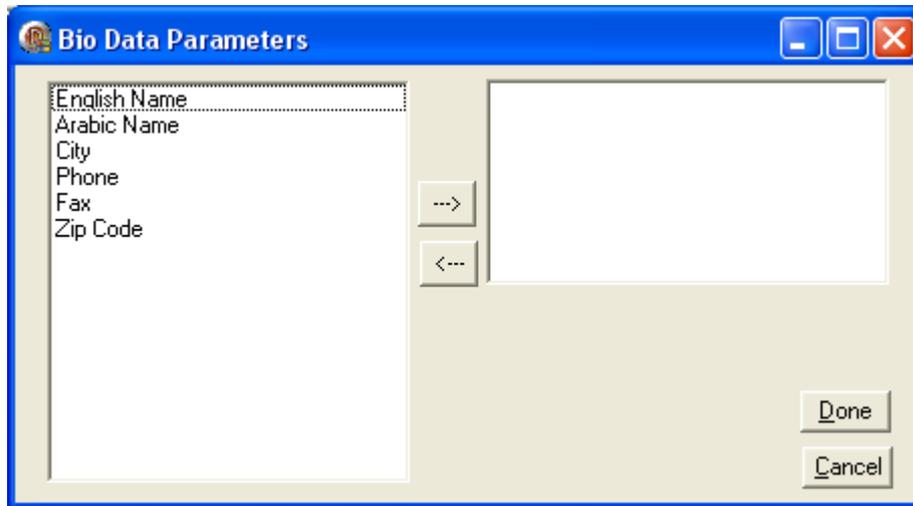
In this dialogue box, the user can select up to five general data parameters. The order of choosing them determines the sort order in the report. Selecting a few parameters, pressing [*Done*], and pressing the [*2. Generate Report*] button on the main screen brings up a screen similar to the one shown in Figure 11 below.

Figure 11: General data report
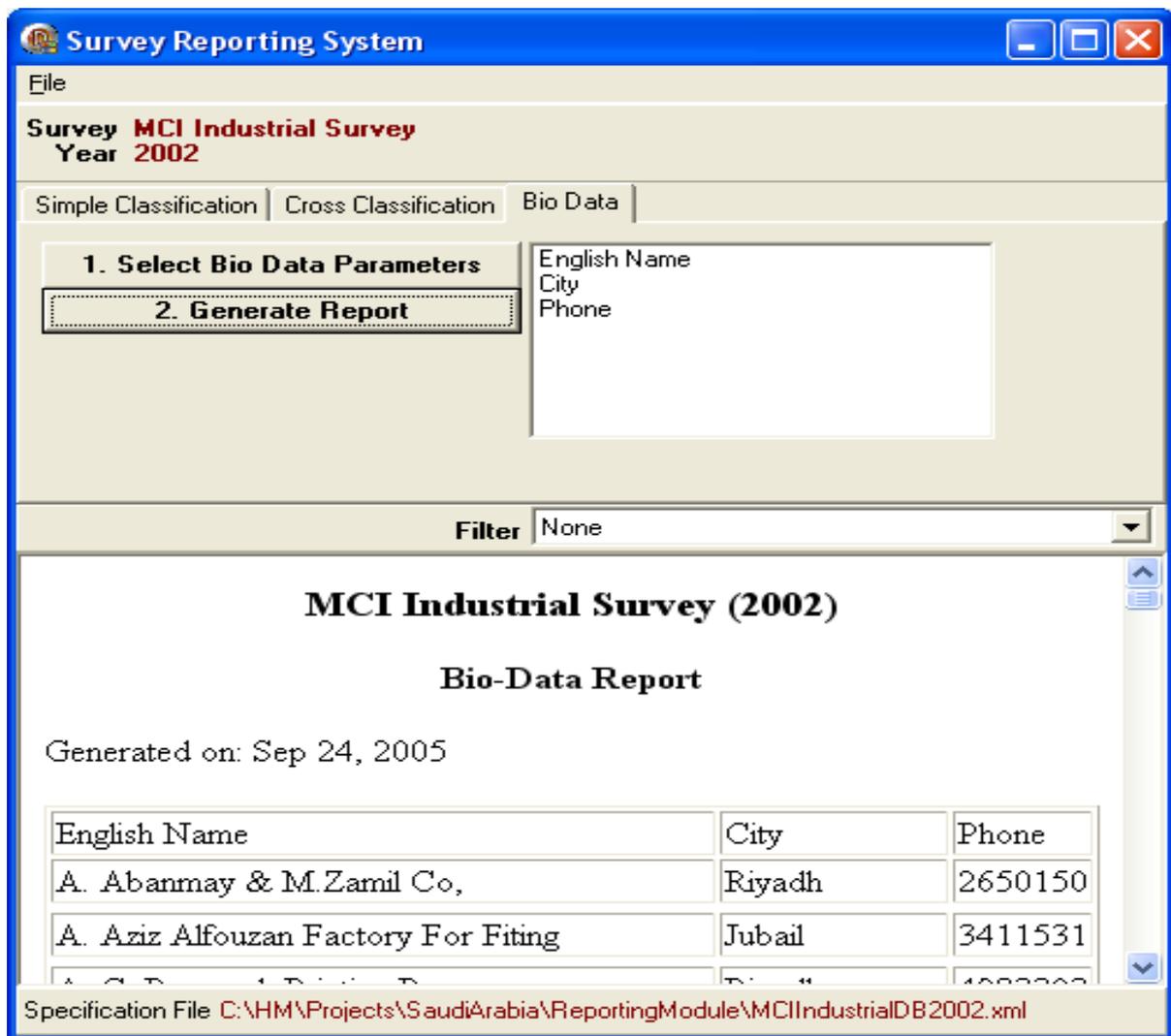
A1.10 Filters

As mentioned in each of the report sections (simple classification, cross classification and general data), an optional step in each of these report generation procedures is the selection of a filter. This filter is a condition (or set of conditions) to which all the records selected for display should conform. A list of filters is defined in the XML specification file.

Figure 12 shows the pull down [*Filter*] menu with 'Production Companies' selected.

Figure 12: Filter
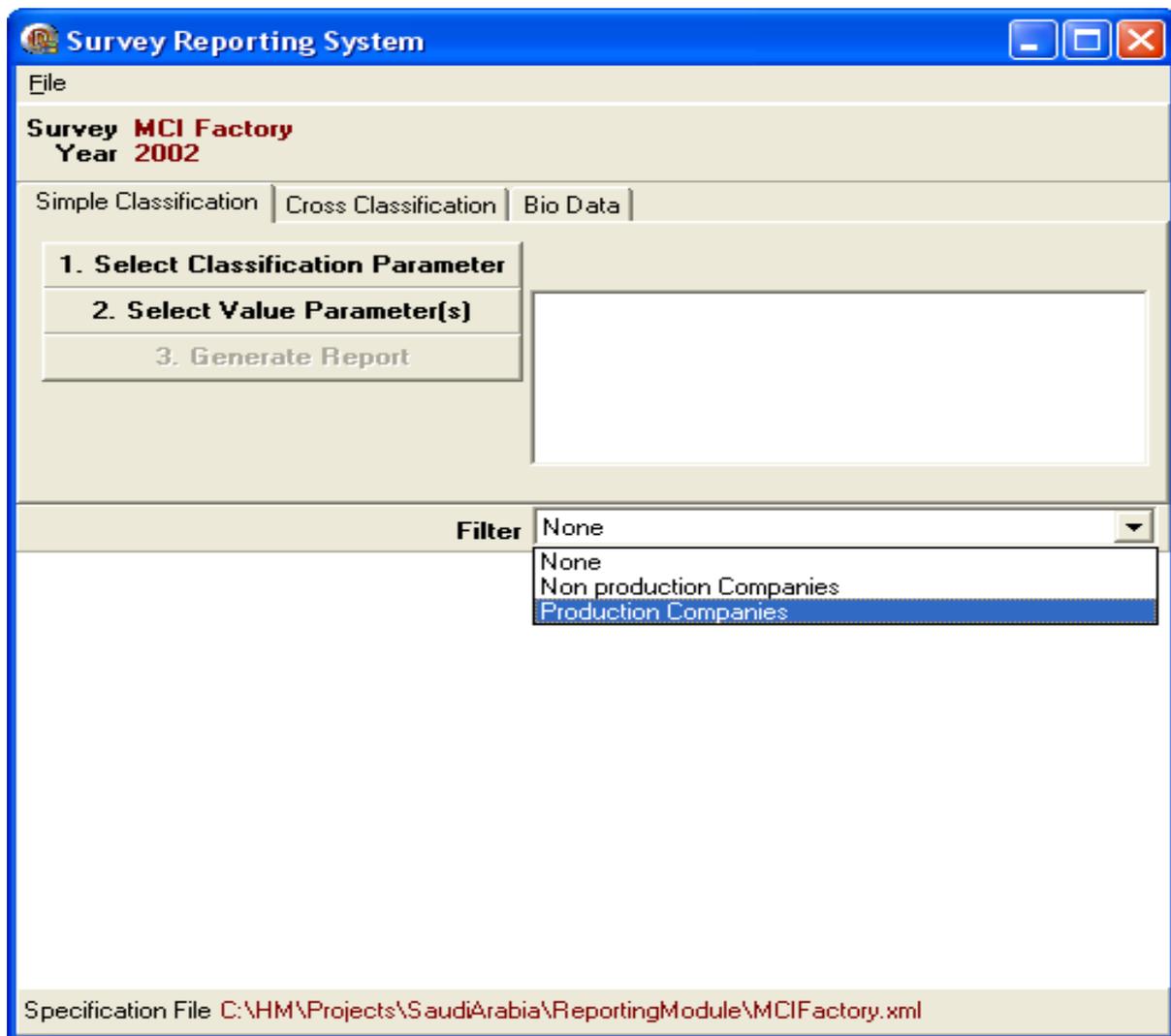
A1.11 Choosing a specification file

As mentioned in the Introduction (A.1.1), the Survey Reporting System is flexible and is driven by a specification file. The user can switch from reporting one survey to another by choosing an appropriately designed specification file in the XML language. This can be done by choosing [*File*]/[*Choose Specification File*] from the main menu. A file selection box comes up from which the user can choose a specification file.

Note that if the specification file does not conform to the requirements of the XML language syntax or if it does not meet the format specification defined in the 'Survey Reporting System-Specification File Format' report, the programme will not work properly and it will signal one or more errors.

A1.12 Saving generated reports in HTML format

Once a report has been generated, it can be viewed and printed as described above. It is also possible to save a copy of the report in HTML format since these reports are already being generated in that format by the system. Under the <install_dir>\output\ directory, the HTML files are generated according to the report type, using the following names:

1) *Simple classification*: **simple.html**

2) *Cross classification*: **cross.html**

3) *General data*: **gen.htm**

If the user wishes to save a copy of the report in HTML file, he/she would need to make a copy of the appropriate report type. Note that the newly generated reports overwrite the old ones so, if an HTML copy is desired, the user should make a copy immediately after generating these reports.

Annex 2:   Sample XML Specification File for Reporting Module

```xml
<?xml version="1.0"?>
<ReportingSystem Survey="MCI Industrial Survey" Year="2001">
   <MainDB Name="MCI Industrial Database"
           MainTable="MST_Company" ODBC="MCIIndustrialDB">
      <Key Field="cmID" Type="number"/>
   </MainDB>
   <LinkedDB LinkTable="Survey_To_Factor_Link"
           MainTable="Factory"/>
   <Parameters>
      <Classification>
         <Parameter Name="English Name" Field="cmNameEnglish"/>
         <Parameter Name="ISIC 3 Digit" Field="ISICCode">
         <Lookup Table="R2ToR3" KeyField="R2"
            LookupField="LEFT(R3,3)"/>
      </Parameter>
         <Parameter Name="ISIC 2 Digit" Field="ISICCode">
         <Lookup Table="R2ToR3" KeyField="R2"
            LookupField="LEFT(R3,2)"/>
      </Parameter>
         <Parameter Name="Legal Type" Field="cmLegalStatus">
         <Lookup Table="MST_LegalStatus" KeyField="Lmlegalcode"
            LookupField="Lmlegalname"/>
      </Parameter>
      </Classification>
      <Value>
         <Parameter Name="Number of Establishments" Table="N/A"
            Field="Count"/>
         <Parameter Name="ISIC-4Digit" Table="N/A"
            Field="LEFT(ISICCode,4)"/>
         <Group Name="Derived Variables" Table="derived">
            <Key Field="id"/>
            <Parameter Name="Gross Output" Field="GO2001"/>
            <Parameter Name="Intermediate Consumption"
            Field="IC2001"/>
            <Parameter Name="Compensation of Employees"
            Field="CE2001"/>
            <Parameter Name="Value Added" Field="VA2001"/>
            <Parameter Name="Gross Operating Surplus"
            Field="GOS2001"/>
         </Group>
         <Group Name="Employment" Table="MST_MANPOWERDATA">
            <Key Field="MMID"/>
            <Parameter Name="Overall Employment"
            Field="Total2001_SaudiNo+Total2001_NonSaudiNo"/>
            <Parameter Name="Emp-Overall Saudi"
            Field="Total2001_SaudiNo"/>
            <Parameter Name="Emp-Overall Non-Saudi"
            Field="Total2001_NonSaudiNo"/>
            <Parameter Name="Emp-Administrative and Clerical-
            Saudi" Field="AdministrativeOther_SaudiNo"/>
            <Parameter Name="Emp-Administrative and Clerical-
            Non-Saudi" Field="AdministrativeOther_NonSaudiNo"/>
            <Parameter Name="Emp-Professional Jobs - Saudi"
            Field="ProfessionalJobsTotal_SaudiNo"/>
            <Parameter Name="Emp-Professional Jobs - Non-Saudi"
            Field="ProfessionalJobsTotal_NonSaudiNo"/>
```

```xml
            <Parameter Name="Emp-Technical and Assistant
            Technical Jobs - Saudi"
            Field="TechnicalAndAssistantTotal_SaudiNo"/>
            <Parameter Name="Emp-Technical and Assistant
            Technical Jobs - Non-Saudi"
            Field="TechnicalAndAssistantTotal_NonSaudiNo"/>
            <Parameter Name="Emp-Other Jobs - Saudi"
            Field="OtherJobsTotal_SaudiNo"/>
            <Parameter Name="Emp-Other Jobs - Non-Saudi"
            Field="OtherJobsTotal_NonSaudiNo"/>
        </Group>
        <Group Name="Wages" Table="MST_MANPOWERDATA">
            <Key Field="MMID"/>
            <Parameter Name="Overall Wages"
            Field="Total2001_Saudi_Total_Annual_Wages+Total2001_
            NonSaudi_Total_Annual_Wages"/>
            <Parameter Name="Wgs-OverallSaudi"
            Field="Total2001_Saudi_Total_Annual_Wages"/>
            <Parameter Name="Wgs-Overall Non-Saudi"
            Field="Total2001_NonSaudi_Total_Annual_Wages"/>
            <Parameter Name="Wgs-Administrative and Clerical-
            Saudi"
            Field="AdministrativeOther_Saudi_Total_Annual_Wages"
            />
            <Parameter Name="Wgs-Administrative and Clerical-
            Non-Saudi"
            Field="AdministrativeOther_NonSaudi_Total_Annual_Wag
            es"/>
            <Parameter Name="Wgs-Professional Jobs - Saudi"
            Field="ProfessionalJobsTotal_Saudi_Total_Annual_Wage
            s"/>
            <Parameter Name="Wgs-Professional Jobs - Non-Saudi"
            Field="ProfessionalJobsTotal_NonSaudi_Total_Annual_W
            ages"/>
            <Parameter Name="Wgs-Technical and Assistant
            Technical Jobs - Saudi"
            Field="TechnicalAndAssistantTotal_Saudi_Total_Annual
            _Wages"/>
            <Parameter Name="Wgs-Technical and Assistant
            Technical Jobs - Non-Saudi"
            Field="TechnicalAndAssistantTotal_NonSaudi_Total_Ann
            ual_Wages"/>
            <Parameter Name="Wgs-Other Jobs - Saudi"
            Field="OtherJobsTotal_Saudi_Total_Annual_Wages"/>
            <Parameter Name="Wgs-Other Jobs - Non-Saudi"
            Field="OtherJobsTotal_NonSaudi_Total_Annual_Wages"/>
        </Group>

        <Parameter Name="Taxes" Table="MST_FINANCIALDATA3"
            Field="licencesandtechnicalassistancefees_2001">
            <Key Field="fmid"/>
        </Parameter>

        <Parameter Name="Raw Materials - Basic and Intermediate
            (Raw Materials Data)" Table="MST_RawMaterialsData"
            Field="rmSaudiOriginValue+rmForeignOriginValue">
            <Key Field="rmID"/>
        </Parameter>

        <Group Name="Balance Sheet" Table="MST_FINANCIALDATA">
            <Key Field="fmid"/>
```

```xml
            <Parameter Name="Goods in Process - Current Year"
            Field="GoodsInProcess_2001"/>
            <Parameter Name="Goods in Process - Previous Year"
            Field="GoodsInProcess_2000"/>
            <Parameter Name="Finished Goods - Current Year"
            Field="FinishedGoods_2001"/>
            <Parameter Name="Finished Goods - Previous Year"
            Field="FinishedGoods_2000"/>
            <Parameter Name="Spare Parts - Current Year"
            Field="SpareParts_2001"/>
            <Parameter Name="Spare Parts - Previous Year"
            Field="SpareParts_2000"/>
            <Parameter Name="Overall Assets"
            Field="TotalFixedAssets_2001"/>
            <Parameter Name="Land - Current Year"
            Field="Land_2001"/>
            <Parameter Name="Land - Previous Year"
            Field="Land_2000"/>
            <Parameter Name="Buildings - Current Year"
            Field="Buildings_2001"/>
            <Parameter Name="Buildings - Previous Year"
            Field="Buildings_2000"/>
            <Parameter Name="Plant and Equipment - Current Year"
            Field="PlantAndEquipment_2001"/>
            <Parameter Name="Plant and Equipment - Previous
            Year" Field="PlantAndEquipment_2000"/>
            <Parameter Name="Motor Vehicles - Current Year"
            Field="MotorVehicles_2001"/>
            <Parameter Name="Motor Vehicles - Previous Year"
            Field="MotorVehicles_2000"/>
            <Parameter Name="Premises Under Construction -
            Current Year"
            Field="PremisesUnderConstruction_2001"/>
            <Parameter Name="Premises Under Construction -
            Previous Year"
            Field="PremisesUnderConstruction_2000"/>
            <Parameter Name="Other Fixed Assets - Current Year"
            Field="FixedOther_2001"/>
            <Parameter Name="Other Fixed Assets - Previous Year"
            Field="FixedOther_2000"/>
        </Group>

    <Group Name="Profit and Loss Account"
        Table="MST_FINANCIALDATA3">
        <Key Field="fmid"/>
<Parameter Name="Sales (P&amp;L)"  Field="sales_2001"/>
<Parameter Name="Raw Materials Used"
        Field="rawmaterialused_2001"/>
<Parameter Name="Rent"  Field="expenditurerent_2001"/>
<Parameter Name="Fuel, Water and Electricity"
        Field="fuelwaterandelectricity_2001"/>
<Parameter Name="Maintenance and Repair Expenses"
        Field="maintenancerepairexpenses_2001"/>
<Parameter Name="Other Direct Expenditure"
        Field="expenditureothers_2001"/>
<Parameter Name="Other Income"
        Field="otherincome_2001"/>
<Parameter Name="Other Rent"  Field="rent_2001"/>
<Parameter Name="Other Expenditure"
        Field="expenditureothers_2001"/>
```

```xml
            <Parameter Name="Selling and Distribution"
                Field="sellinganddistribution_2001"/>
            <Parameter Name="Other Expenditure"
                Field="otherexpenditure_2001"/>
            </Group>
            <Parameter Name="Sales B (Sales Data)"
                Table="MST_SalesData" Field="smSalesValue">
                <Key Field="smID"/>
            </Parameter>
            <Parameter Name="Exports" Table="MST_ExportData"
                Field="emValueOfExportSales">
                <Key Field="emID"/>
            </Parameter>
            <Group Name="Public Utilities B (Public Utilities
                Data)" Table="MST_PUBLICUTILITIESDATA">
                <Key Field="cmid"/>
                <Parameter Name="Fuel"
                Field="iif(Srno=8,Cost2001,0)"/>
                <Parameter Name="Electricity"
                Field="iif(Srno=13,Cost2001,0)"/>
                <Parameter Name="Water"
                Field="iif(Srno=18,Cost2001,0)"/>
            </Group>
            <Parameter Name="Raw Materials - Packing/Packaging"
                Table="MST_PackingMaterialsData"
                Field="pmSaudiOriginValue+pmForeignOriginValue">
                <Key Field="pmID"/>
            </Parameter>
    </Value>
      <BioData>
            <Parameter Name="English Name" Field="cmNameEnglish"/>
            <Parameter Name="Arabic Name" Field="cmNameArabic"/>
            <Parameter Name="City" Field="cmCityCode">
                <Lookup Table="MST_City" KeyField="cmCityCode"
                LookupField="cmCityName"/>
            </Parameter>
            <Parameter Name="Phone" Field="cmTelephone"/>
            <Parameter Name="Fax" Field="cmFactoryFax"/>
            <Parameter Name="Zip Code" Field="cmZipCode"/>
    </BioData>
   </Parameters>
</ReportingSystem>
```

Chapter 5: An index system to measure short-term and medium-term changes in industrial activities

1.0   Introduction

One of the major uses of industrial statistics is to support the compilation of annual National Accounts. Even when annual industrial and other Economic Surveys are conducted, the information collected through them becomes available only some time after the end of the reference period. Similarly, when company annual accounts are used for the corporate sector, there is a substantial time lag in their availability.

For effective management of the economy, as well as policy formulation, it is necessary to have information available as early as possible. The initial estimates of the level of economic activity should be available by the end of the same year. Furthermore, it is becoming regular practice for countries that are still developing their statistical system to compile quarterly national accounts. For both annual and quarterly national accounts, the use of *"short-term indicators"* for the measurement of changes in levels of economic activities is necessary.

Short-term indicators come in many guises and their commonality lies in the assumption that a particular time series gives a good indication of the change in the level of activity (volume or value of outputs) and/or prices in the activity concerned. In general, the assumption of representativeness of an indicator should be checked from time to time as, in many cases, relationships between indicators and activities do change over time.

For industrial activities, (herein defined as ISIC Sections B. Mining and quarrying; C. Manufacturing; D. Electricity, gas, steam and air conditioning supply and E. Water supply, sewerage, waste management and re-mediation activities[1]), the relationship between indicators and production is generally strong but for some other activities good indicators can also be found. The room occupancy rate in the hotel industry; the number of call minutes for telecommunications; and the volume of letters and parcels for the post office are some examples.

The regular measurement of changes in production volumes and prices in these activities is done most easily with the help of a system of production and producer price indices. In these guidelines, it is assumed that coverage is restricted to industrial activities as commonly defined.

---

[1] Sections B to E of the International Standard Industrial Classification of All Economic Activities (ISIC), Rev. 4.

Production indices measure the changes in volumes of production, while the producer price indices measure the change in price levels and so provide information for the conversion between constant and current prices. The simultaneous compilation of both indices not only reduces the response burden and simplifies the data collection, processing and compilation processes but also enhances the value of both series.

It is not usual to find integrated systems for these index series, largely because of the way activities have traditionally been distributed in statistics offices[2]. Industrial statistics and price statistics are generally collected by different sections of the office, even where the respondents are the same. Again, the discussion here is basically limited to production indices, as these fall within UNIDO's remit. In general, the description of producer price indices is the same as that of production indices and, where this is not the case, the differences are noted.

The strength of an index system lies in its ability to measure changes with a minimum of data frequently collected. These data are linked to the prevailing structure through a system of weights. Under normal circumstances, this provides accurate estimates of the developments over short- to medium-length time periods. Structures generally change over time and the base-weights system for an index should be revised regularly in order to incorporate such changes. It is generally recommended that re-basing should take place at five year intervals.

In countries with a VAT system, however, there is now another method for keeping the index weights up to date. The value added tax system can provide a continuous, annual set of weights that then automatically reflects the changes in the production structure. This would allow the use of a chained index series to measure industrial performance and will take into account structural changes as they occur. It would also be possible to calculate production indices such as a Laspeyres series and producer price indices using Paasche methodology and then generate overall Fisher indices as the average of both.

2. Methodology

2.1 General

The methodology for the preparation of any index is straightforward. In short, it can be described as a weighted set of ratios between the current period and the selected base (or

---

[2] In terms of recommendations and guidelines for industrial production and producer price information, a similar dichotomy exists. UNIDO is the international agency responsible for industrial statistics while producer price methodology is the domain of the IMF, OECD and various other agencies.

reference) period. This is the same for a production index series, a price index series, or other series such as unit values of imports and exports.

Data collection for an index series consists of two distinct activities:

- First, there is an initial collection of some baseline and back-period information, which needs to be repeated every five years or so;

- Second, there is the actual (monthly or quarterly) data collection for the index calculation.

Where feasible, it is always desirable to base an index series to the new base year for the national accounts. As that is generally a few years earlier, this means that price and production information for the period following that newly selected base-year needs to be collected from the selected companies during the initial introductory round. After this exercise is completed, the regular monthly data collection for the index calculation can begin. It may be noted that, even when the data are mainly used for the compilation of quarterly national accounts, a monthly collection cycle is preferred over a quarterly one. As not all respondents provide information on time, missing information in a quarterly series is much more problematic than in a monthly series as no information is available for those units in the first case, while there may still be two months of data available in the latter case.

2.1.1. Commodity and industry weighting

The two main methods used to calculate a production index can be categorised as commodity and industry weighted. The index calculations compiled by many agencies are commodity-weighted. However, the index proposed here is of the more accurate industry-weighted type.

The difference between the two methods is that the commodity-weighted procedure assumes that the selected commodities accurately reflect their relative importance in industrial production, something only rarely true and then only by chance. The main reason for this is that the commodities generally used in the calculation of the index are not selected for that purpose. It is simply that information on their production happens to be available from administrative records, often because they are subject to price controls, excise duties or other taxes.

The industry-weighted procedure starts off with the same data but has weights at different levels of aggregation, thereby assuming that the commodities, enterprises and industries covered in the index are representative of the enterprises, industries and industry groups of

which they are part. This is a more reliable assumption than the one implied by the commodity-weighted method.

2.1.2. Selection of respondents

In most branches of industry, there are some (very) large companies, some medium-sized and a number of small ones. It is therefore much more efficient to have a purposive sample than to use random sampling to obtain the list of respondents.

A 70/70/70 rule is useful in assisting the selection of respondents for the survey. There are three steps in the selection process for sorting the enterprises or establishments by activity and size, both in descending order, as follows:

1) Select the largest activities (at 4-digit ISIC) which, together, account for at least 70 per cent of the total

2) Within each of the selected activities, select the largest enterprises which account for not less than 70 per cent of the total in those activities

3) For each of the selected enterprises, select product lines totalling 70 per cent of the production

In the first two steps, size is measured by value added at basic prices (excluding taxes and subsidies) while, in the third step, the output values are selected. The latter is done because the value added is normally not available for individual product lines within enterprises.

In general, this procedure results in the selection of between 150 and 200 enterprises, with some 250 to 300 product lines in total. There will be a few activities with no large operations but many smaller participants. For those activities, the 70 per cent rule can be relaxed. Provided that no less than ten units for those activities are in the selection, the changes in output will be   adequately registered.

It may be noted that no attention is paid in this selection procedure to small activities and units operating within them. This group may, in principle, have different rates of change than the larger units and activities. However, so long as the weights system is updated on a regular basis - at least once in a period of five years - any distortion caused by the selection procedure will only have a marginal effect. For example, in terms of employment, a doubling of size in a one-man unit is the same as a one per cent change in a unit employing 100 workers. Thus, even significant differences in developments in the short run cannot have large effects on the total.

The main arguments against inclusion of small units in the sample are efficiency and costs. Large units can be reached, after proper introduction, by various means of communication that do not require regular visits to the plants or offices of the enterprises. Information can

be submitted by mail, e-mail, fax, telephone or even logged directly into the database (as shoppers on the internet do). The smallest units cannot be reached easily and field visits are expensive and time-consuming. Thus, for marginal improvements, at most, in the accuracy of the results, major costs would have to be incurred.

2.1.3.   Quantity, value and volume measures

The output of an enterprise can be measured in different ways. It could be measured as a quantity, that is the physical measurement of output such as the quantity of cement or the number of cars produced. It could also be measured as the volume which would, in the latter case, count the number of cars by type rather than in total. For that reason, volume measures tend to be given in monetary values and not physical units of measurement. Output could also be measured in monetary values, that is, the value of sales corrected for changes in stocks of finished products. There are advantages and disadvantages in the use of any of these measures and the choice largely depends on the activity covered and, to a lesser extent, on the situation on the ground.

Quantity

For homogeneous products, the quantity of production would be the simplest, and probably also the most accurate, measure of output. This applies, for instance, to cement factories, grain and sugar mills, steel re-rolling units, and many activities on a smaller scale with a limited product range. Basically, when the unit values of different varieties or qualities of the products manufactured are similar, the quantity measure will give a good indicator of production.

Industrial production statistics often also give the total production of the most important goods, such as the number of cars, the tonnage of cement and sugar. The data gathered through the production index system are not geared towards the preparation of these types of summary totals because not all establishments producing the selected goods are included in an index system. For the generation of these kinds of statistics, standard units of measurement across the enterprises (and for international comparison in international standard units) would be necessary. In this case, selection of the most convenient reporting unit, as suggested in the previous paragraph, would not be possible.

The comparison of total production volumes over time, even for homogeneous products, does not give an accurate measure of changes in production in national accounting terms. This procedure should therefore be avoided. An example can illustrate the problem. In cement manufacturing, the traditional "wet" process is very energy-intensive. The more recent "dry" process is much more energy-efficient and therefore generates a higher value added per unit output than the older technology. However, it is rather expensive to convert an old unit to the new technology which means that new capacity alone uses the new technology and old capacity retains the old technology for the lifetime of the unit. Expansion of capacity and production in this case therefore will result in higher value-added

ratios over time (as new technology is used for new capacity). When the production of individual units is measured and compared over time, this problem does not exist as the production of each is evaluated against its own value added ratios.

It is, in many cases, also possible to measure output on the basis of the quantity of major inputs used. Depending on the type of activity, there may be a strong relationship between the quantity of the major inputs and the volume of outputs produced. In various industries, the amount of materials used also largely determines the costs of the final product. For example, jewellery is generally costed on the basis of weight and the current price of gold although sometimes extra costing is done when more elaborate processing is involved. The quantity of paper used may be a good indicator for printing and the quantity of steel used for metal casting.

Volume

For activities where a variety of qualities and sizes of the same or similar goods are produced, simply adding the quantities will not give a proper indicator of output when unit-values differ significantly. It may be expected also that the mix of qualities will change over time, thereby changing the output per unit of quantity. Hence, a more accurate measure is required. The volume of output is such a measure. It may be described as a weighted average of the quantities of output using some fixed indicator as a weight. This may be the price (in the base period) but also the cost of production or some other indicator that the enterprises use for measuring and aggregating their own output.

For practical purposes, volumes are often expressed in monetary values but an industry may have its own measure. In shipping, the worldwide standard unit is the 20 foot container and all other sizes are expressed in equivalents. Thus, a 40 foot container is two standard units while two half-height 20 foot containers make one standard unit. Similarly, local or international standards exist in other industries. It is reporting convenience that should govern the selection of unit of measurement as index calculations are done with the same efficiency whatever unit is used. Provided that the nominator- giving current production - and the denominator - giving the production in the base period - use the same unit of measurement, the resulting ratio has no dimension at all. Therefore, the unit used by any of the respondents is not relevant for the others and, in principle, the unit of measurement used by each of them could be different.

Value

The measurement of output in values of production is common for activities with a wide variety of products or qualities. For example, it is a widely used method in printing and furniture. The production values are then deflated with an appropriate price index to obtain constant-price values. The price index series most commonly used for deflation, if available, is the producer price index.

The International Recommendations for Index Numbers of Industrial Production 2010 give preference to the deflation method. However, there are some difficulties in its implementation in developing countries. One reason is that the quantity or volume information has a wider range of use and, therefore, has value in its own right. In addition, price index series in many developing countries are not so reliable.

Independent estimates for volume and price changes are therefore preferable. Furthermore, with two independent series, it is possible to check their combined accuracy. By definition, current production values in current prices should equal the previous year's values adjusted for both volume and price changes. Statistically, this relationship is:

$$Y_t = Y_{t-1} \cdot \frac{I_{p_t}}{I_{p_{t-1}}} \cdot \frac{I_{q_t}}{I_{q_{t-1}}}$$

where:

Y = output

I_p = price index

I_q = volume index

t = current period

$t-1$ = previous period

When no independent volume information is available, the consistency of price and volume change cannot be checked and errors in the price index series will inevitably lead to distortions in the estimates for production volumes. This, in turn, will affect the most important GDP estimates, namely those from the production side at constant prices

2.1.4. Preparation of weights sets

The weights sets form the key element of any index system. They are used to create aggregates of the information in a variety of formats. There may be different types of indices calculated, such as Laspeyres, chained Laspeyres, or Paasche. For each of these types of indices, the production ratios at product level for each individual product and company are the same, as are the indices at the lowest level of aggregation, the basic indices, for each product. These indices are subsequently grouped into higher-level aggregates, culminating, after various steps, in an overall index.

The grouping of the low-level indices can be done in various ways with the one chosen depending on a number of factors, such as past practices, or the regionally or internationally recommended classifications. Occasionally, more than one set of indices is produced in order to meet the demands of different users but this is not very usual. The most common groupings are based on the International Standard Industrial Classification of All Economic Activities, revision 4 and the Central Product Classification (CPC). Nonetheless, a large number of countries continue to use their own traditional groupings[3].

The aggregate indices at each level are calculated as weighted averages of the indices at the previous level. The weights are derived from Industrial Surveys or the VAT system database, if available. For production indices, the value added is used to calculate the weights while, for producer price indices, gross output is used.

There may, on occasion, be problems in determining appropriate weights for some activities, as value added for these may be negative or far too small for the importance of the activity in the economy. This happens when sales prices are controlled and subsidies that were promised were either paid late or not paid in the form of product subsidies but as capital grants or other forms of support. For example, petroleum refining and electricity production are activities that suffer from these problems, especially when input prices are increasing very fast. In such cases, the value added could be imputed on the basis of the contracted or promised compensation. This could be a subsidy per unit of production or a certain percentage return on investment. The latter is a common condition in operating contracts for large-scale projects such as oil refineries.

After the imputed value added for the affected units (if any) has been substituted for the original information contained in the database, all aggregates are recalculated using these imputed values. The weights for the activities and enterprises are subsequently calculated. The weights for the product lines within the enterprises can only be calculated using additional information, either from the Industrial Survey, if one has been compiled, or from the special baseline survey that forms the starting point for the establishment or revamp of the production index series. Data requirements and collection for the baseline information are described in the following section.

2.1.5.   Extension to Chained Indices

In countries where a VAT system exists and details from that source are available to the statistics office, it will be possible from the start of the index compilation to calculate more than one set of base-year weights. In these countries, it is therefore possible to develop index series on the basis of current weights (Paasche) and also chained indexes. A series based on the chained index is probably the most useful for the production indices in

---

[3] The PRIMA software described in Annex E allows for the parallel calculation of as many index series as there are weight sets available.

developing countries where the production structure is changing relatively fast. This is because a chained index takes into account the changes in the industrial structure of the economy on a continuous basis.

The use of a chained index based on VAT information removes the main drawback of the standard Laspeyres series for use over long periods, namely that it cannot accommodate changes in the structure. It also reduces the need for frequent large-scale Industrial Surveys or censuses since the main benefit of these data sets in comparison with VAT data is the detailed information on the production structure for each activity that can be derived from the survey information. The VAT information gives only a good approximation of the broad measure of value added, without details of the input structure and composition of value added.

VAT data contain, by definition, information on the value added generated by each company, as well as on the total turnover of the company. Depending on the underlying legislation, there are smaller or larger differences between the value added derived from the VAT data and that defined in the national accounts. This only matters, however, if the difference between the coverage of value added in the two definitions varies significantly from company to company or from activity to activity. If there is a consistent difference between the two, both between activities and over time, the VAT measure can be used without further adjustments as, in this case, relative sizes are reflected properly. Otherwise, it may be necessary to use the ratios between VAT value added and GDP value added for different activities in a year for which industrial census data are available in order to derive adjustment factors.

When VAT information is available, the compilation of a producer price index can best be in the Paasche form. In this index type, the price changes registered during each year are weighted by the production structure of the same year. Combined with a Laspeyres production index, the Fisher "ideal" index can also be calculated in these cases.

2.2. Baseline information

The base weights of the index system describe the relative importance of the particular commodity or industry in the overall industrial activity in the country. For production, this importance is usually measured l not by gross output but by value added. The latter is the statistical concept which basically measures the income derived from the activity by the providers of capital (owners and lenders) and labour (hired labourers and owners/unpaid family workers). For analytical purposes, the value added should be measured at basic values, that is, it should not include indirect taxes and subsidies, as these are transfers to

and from government rather than income for the enterprise (for example, a cigarette factory is a *de facto* tax collector).

However, producer price indices should be based on the share in gross output of the establishment (normally also measured in basic prices), since price changes affect both the total (gross) value of output and value added.

The weights for the index series are commonly derived from the (annual) Industrial Survey. For each industry, current production and price data are sought for all establishments with a share of output above a cut-off point (for example, of two per cent) in their respective industries. The weights for the index series can also be derived from the data available from the VAT service in countries with a VAT system and infrequent Industrial Surveys.

In order to simplify baseline data collection and re-basing for an index system weights for both the production and price index systems can also be based on the value of sales (approximating gross output) as this information is often easier to collect. Nonetheless, value added (net output) would be the more appropriate weighting system for production indices, if this information can be obtained.

Thus, for the base year, it is necessary to obtain information on the value of production (or sales) for all companies in the manufacturing sector; or at least those in the formal part of the sector. The share of a company's sales in the sector forms its weight in the sector. Changes in the volume of production from month to month are then sufficient to calculate the average change in output of the sector. A sample baseline and back-period data questionnaire is shown in Annex 2. It may be noted that, in many countries, the output or sales values can also be provided by the VAT service, in which case the same information requested on the form serves as a cross-check to ensure that the data provided relate to the entire company rather than only to some part of it.

In addition to this basic information, it is also necessary in the initial questionnaire to collect information on the most appropriate product line(s), item groups and pricing items, along with their respective units of measurement, since this needs to be used in the subsequent monthly surveys. As a common rule, only "important" product lines are covered in the survey. Within these, only the most common items or item groups are selected.

For a pricing item, the most common packaging and size is used. The general idea is that the item(s) selected should be representative of the production of the company, a choice that can best be made by the company itself. Therefore, the initial questionnaire should ask for this information.

For both the base year and the current periods, information on the quantity of production (of the most important item, as indicated earlier) and the price of a standard unit (that is,

most common packaging and size) of the (most common) pricing item is sought. For general and statistical use, information on employment and the number of days and shifts worked is also collected.

Information on production is collected for the major line of products of the establishment (e.g. the output of wheat flour or palm oil) and, where more lines are of similar importance, more than one production series is included (e.g. the output of milk and ice cream for a dairy plant).

For the producer price index, the most important product of the major line of an establishment is used as the pricing item. The reference is very detailed to ensure consistent reporting (e.g. a case of 24 glass bottles of 300 ml Coca Cola; one metric tonne of crude palm oil). In cases where an establishment reports more than one line of product, a pricing item is selected for each of those.

2.3. Recurrent data collection

The indices are computed in a number of steps. The underlying assumption is that the production and price reports for a particular product line are representative of the establishment. Similarly, it is assumed that the establishments within an industry (four-digit ISIC) are representative of that industry, and so on. This representativeness is guaranteed by including the largest units from each industry in the index system. In general, it is advisable, where possible, to include units accounting for more than 70 per cent of the industry's production. However, in the more widely distributed industries (e.g. textiles, food products), a lower percentage can be used as the production processes from unit to unit are similar although it is still the largest units which are selected. Weighting takes place at each level of aggregation.

The weights are representative of the entire activity rather than the units covered in the index. The implicit assumption in the system is, therefore, that the product lines, establishments or industries not enumerated behave in a similar way to the average of all reporting units or groups at that level of aggregation. It also means that an automatic adjustment for non-response is in place, using the same assumptions, namely that the behaviour of non-respondents is the same as the average behaviour of the respondents.

As a first level of aggregation, if more than one product line is reported by an establishment, the production and price changes for the different lines are averaged to obtain the indexes for the establishment. The averages at this level are not weighted.

In cases where the indices are aggregated on the basis of ISIC, the following scheme could be used.

- The production and price changes for all establishments within a four-digit ISIC group are averaged on the basis of the relative share of these establishments within the industry.

- The production and price changes measured at the four-digit ISIC industry level are subsequently averaged to obtain the 3-digit ISIC major group indices. Again, the averages are weighted, this time with the relative share of the four-digit industry group in the three-digit major group.

- The three-digit major group averages are then weighted to obtain the two-digit ISIC division averages. This is the level of detail that is normally published.

- As the last stage, the ISIC divisions are weighted to obtain the overall manufacturing indices.

A simplified set of two alternative calculations of the industry-weighted production index system (with or without detailed enterprise data) is shown in Annex 1.

As there tend to be changes in product lines, production patterns and most common packaging and sizes over time, the monthly questionnaire needs to keep track of these, where practicable. The most common change is with respect to pricing items, as packaging styles, materials and sizes tend to change over time. In these cases it is necessary to "chain" the series so that price changes continue to be reflected as accurately as possible.

2.4. Data Collection and Processing

The timeliness and accuracy of data collection is a major determinant of the quality of the information. As a starting point in the launch of a new system, it is important that the entrepreneurs are made aware of the activity and its importance. It would be very useful if a timeslot could be secured by the Statistics Office at a general meeting of the major employers associations to introduce the new system. The general introduction to the new producer price and production index system can be used as a guide for a presentation at employers associations or other industry associations.

Initially, all the selected units should be visited by experienced[4] statistics office staff to obtain the data and to explain the purpose of the data collection. Such visits are repeated for as long as necessary on a monthly basis to collect new data but, at the same time, companies are encouraged to report every month by mail, fax or e-mail in a more or less automatic fashion.

---

[4] Sending junior staff to meet company management is counterproductive as they cannot respond to the specific questions raised by managers. Since these visits involve a fairly limited number of companies, they should be conducted by the head of the responsible unit to the greatest extent possible.

The ultimate aim is for all information to be transmitted regularly by one of these means without the need for visits. This both minimises the costs of data collection and can improve its timeliness. For respondents, the costs are also minimised as the reports can be filed at their convenience (for example, as part of the monthly reporting system in the company) while, in a system designed around field work, they need to attend to visitors to their premises and this can often occur at an inconvenient time for the entrepreneur. A sample monthly questionnaire (and its pre-filled complement) is presented in Annex 3.

The index calculations are based on a comparison of the quantity of production over time. This can only been done when the mathematical result is without a dimension, that is, when the unit of measurement in both periods is the same. For instance, the change in production of cement by company A is compared between the current period and a previous period. If the quantity is given in tonnes in one period and sacks or kilograms in the other period, a comparison cannot be made so the quantity in either of the two periods first needs to be expressed in the other unit. In this case, a conversion from kilos to tonnes will be straightforward but, without extra information (on the contents of a standard sack), a conversion between sacks and tonnes is not possible. The importance of reporting information using identical units of measurement over time cannot, therefore, be overemphasised.

It may be assumed that, at the time of calculating the index, not all the company information will have been obtained. Therefore, it is necessary to impute a value for missing information. In the system described above, this is done automatically at all levels of aggregation. The denominators of the formulas used at each stage include only the weights of units or industries also included in the numerators. The implicit basic assumption of this procedure is that the rate of change for the reporting and non-reporting units has been the same. Without evidence to the contrary, this is the most logical assumption to make. Care should be taken, however, to ensure that all those units which did not have any output during the period (e.g. seasonal units) have been included with a zero production report rather than have been omitted.

The index series for the measurement of industrial production should be designed as an integral system that will produce base-weighted Laspeyres indices, current-weighted Paasche indices and chained indices.

2.5. Additional questions

If more analytical information on input structures is required, some additional questions can be included, if practicable, in the baseline survey questionnaire. The type of question to be asked (all expressed in terms of percentage of sales or output) could be:

- Percentage share of imported inputs in value of sales

- Percentage share of local inputs in value of sales

- Percentage share of costs of utilities in value of sales (split between electricity, telephone, water)

It should be noted that the set of sample survey questionnaires presented at the end of this chapter do not include these last questions.

3.   Publication of results

The production and producer price indices should be published on a monthly basis in the form of a standard series of tables. A sample set of tabulations is given below. Once information becomes available for multiple base-weight periods, the series of tables will be repeated for the Laspeyres, Paasche, and chained indices. The information should also be made available electronically on the Statistics Office website. Using the PRIMA reporting module, this can be extended into an interactive procedure whereby the user can design customised sub-series as may be required for analytical purposes.

3.1. List of tables for production indices

1.1   Monthly production indices by industry division for the last 13 months

1.2   Percentage change from the previous month in production indices by industry division for the past 13 months

1.3   Summary of average production indices by industry division and year

1.4   Summary of average production indices by month and year

1.5   Production indices for the current year by month and industry division, calculated as the average index for the year-to-date for the month reported

1.6   Percentage change in the production indexes for the current fiscal year as compared with the level in the same period during the previous year

3.2. List of tables for producer price indices

2.1   Monthly producer price indices by industry division for the last 13 months

2.2   Percentage change from the previous month in producer price indices by industry division for the past 13 months

4.0 Limitations of the index series

The major limitation of the series at this stage of development is the limited coverage, with only a single, or a few, reporting units for the smaller industry divisions. Thus, while the overall index and these sub-indices for larger divisions, such as food, will give a fairly accurate picture of the development of production over time, the indices for smaller industry divisions will, from time to time, show large fluctuations due to the limited coverage.

A second weakness is that only a small base exists, so far, for many important industries. In other words, the establishment of a single new large-scale company may result in large changes in an industry (e.g. cement, cigarettes, petroleum). A solution to this problem is not so straightforward although a judicious use of a mix of methodologies, by calculating the various types of indices simultaneously, can improve the situation. The system provides for the calculation of Laspeyres (base-weighted) indices, as well as Paasche (current weighted), and chained indices.

5.0   Possible further developments

The production and producer price index series proposed in these Guidelines can be expanded and improved upon in various ways. An extension to chained indices and the introduction of regional index series are briefly enumerated below. Measurement of employment trends and seasonal production patterns are other possible additions. Other important indicators that may be considered, such as investment and environmental impact, would require separate data collection efforts as the frame and procedures of the production index series will not yield useful results in these fields.

5.1 Extension to Chained Indices

When more than one set of base-year weights has become available, it is possible to develop index series on the basis of current weights (Paasche) and also chained indices. While these have several shortcomings of their own, in combination with the Laspeyres index series initially compiled, they provide additional insights into the development of industry. Their compilation can be carried out as soon as these new base-weights become available since no additional data collection is required.

5.2 Extension to Regional Index Series

As the identification codes of establishments used in the system are based on location, it is rather simple to develop a series of indices on the basis of regional distribution although this is not a first priority. Once the system has been established, its further development is straightforward.

Annex 1: Production index: calculations of the weights

A1.1 Option 1 – detailed data by enterprise are available

Weights based on:	Base year 2002		Industry weights Value Added		Enterprise weights Value Added		Product weights Gross output	
	Gross output	Value added	Overall	Covered	Overall	Covered	Overall	Covered
	A	B	C	D	E	F	G	H
Industry 1	**10000**	**3000**	0.3158	0.3529	**1.0000**	**1.0000**		
Enterprise 1	1000	250			0.0833	0.0962	**1.0000**	**1.0000**
Product A	*500*						*0.5000*	*0.5556*
Product B	*400*						*0.4000*	*0.4444*
Other products	*100*						*0.1000*	*--*
Enterprise 2	2000	650			0.2167	0.2500	**1.0000**	**1.0000**
Product B	*500*						*0.2500*	*0.2941*
Product C	*1200*						*0.6000*	*0.7059*
Other products	*300*						*0.1500*	*--*
Enterprise 3	1500	400			0.1333	0.1538	**1.0000**	**1.0000**
Product A	*500*						*0.3333*	*0.3333*
Product C	*1000*						*0.6667*	*0.6667*
Enterprise 4	4000	1300			0.4333	0.5000	**1.0000**	**1.0000**
Product A	*3000*						*0.7500*	*1.0000*
Other products	*1000*						*0.2500*	*--*
Other enterprises	1500	400			0.1333	--		
Industry 2	**12000**	**3000**	0.3158	0.3529				
etc…								
Industry 3	**8000**	**2500**	0.2632	0.2941				
etc…								
Other Industries	**3000**	**1000**	0.1053	--				
Total Industry	**33000**	**9500**	**1.0000**	**1.0000**				

Note: For each level the weights in column "**Covered**" (*columns D, F and H*) are used.

A1.2 Option 2 – no data by enterprise are available; only product output values

Weights based on:	Base year 2002		Industry weights Value Added		Product weights Gross output	
	Gross output	Value added	Overall	Covered	Overall	Covered
	A	B	C	D	G	H
Industry 1	**10000**	**3000**	0.3158	0.3529	**1.0000**	**1.0000**
Product A	4400				0.4400	0.5432
Product B	1200				0.1200	0.1481
Product C	2500				0.2500	0.3086
Other products	1900				0.1900	--
Industry 2	**12000**	**3000**	0.3158	0.3529		
etc…						
Industry 3	**8000**	**2500**	0.2632	0.2941		
etc…						
Other Industries	**3000**	**1000**	0.1053	--		
Total Industry	**33000**	**9500**	**1.0000**	**1.0000**		

Note: For each level the weights in column "**Covered**" (*columns D and H*) are used.

Annex 2:   Instructions for completing the baseline survey questionnaire

The baseline questionnaire for the new Product Price and Production Index System is used to link the system to the base period of the National Accounts series, indicated in the following as year *yyyy*. Generally, it takes some time to decide which year should be used as the new base period for the national accounts. The general state of the economy in the planned new base year should be as "normal" as possible. Some criteria for a "normal period" include: i) no extremely high or low prices for either the major exports or imports; (ii) a level of inflation similar to that of the surrounding years; and (iii) no major droughts or floods or other disturbances to agricultural production.

Each of the companies selected for the survey is requested to fill out the questionnaire shown below. Up to three sets of the questionnaire are provided for each production location of the company, that is, one for each important product line at the location. An important product line is thereby defined as one that has accounted for more than one quarter of the value of output at the location.

In order to effectively manage communications for the survey, which will be mail-based (regular mail, fax, e-mail) and, at a later stage, direct data entry on the Statistics Office website and to reduce the burden on respondents, the relevant contact information is requested in the first section of the questionnaire. This section also includes a question on the total value of sales from own production at the location (excluding any taxes) during year *yyyy*. This information is used to calculate the base period weights.

In the second section, a description of the selected product line is requested with one form being filled out for each product line which accounts for more than one quarter of production. The description of the product generally follows the product line but it is repeated here as there may be instances where there are more products coming from the same line. If this is the case, the most important product should be selected for reporting.

The unit of measurement in which all quantities are reported should be the one most convenient for the product line; that is, the one used internally by the company.

For the years *yyyy* and *yyyy+1*, annual production (measured in the unit described above) is requested. For the year yyyy+2, the same information is requested on a monthly basis.

In the third section, information is sought on the price of the most common product of the line in the most widely used packaging. In the example given in this data set, the Coca Cola company is assumed to have two main product lines, namely soft drinks and water. The production of either can be measured in hectolitres, with the distinction of the various flavours of the soft drinks being ignored. However, for consistency in pricing, the distinction is important. Therefore, if the most common item sold during the year was Coca Cola, in cases of 24 bottles of 300 ml, this is the appropriate pricing item. The price to be reported is the one that is most commonly charged during the period. Thus, if a 10 per cent discount was generally granted on the list price, this discounted price should be reported.

For the years *yyyy* and *yyyy+1*, the year end price (for the item and in the unit described above) is requested. For the year yyyy+2, the same information is requested on a monthly basis.

In the fourth and final section, a few characteristics are requested for the product line. This will enable the Statistics Office to calculate productivity changes and to normalise production. The three items are:

1) The number of employees at the end of the year *yyyy*

2) The number of working days during the year

3) The number of shifts worked per day. This can be indicated as one, two or three shifts per day when these have been standard throughout the period. Otherwise, the total number of shifts worked during the year should be reported. For example, if there were 320 working days and operations during 160 days were on a single shift basis, while the remainder was on double shifts, the total number of shifts worked is 160 x 1 + 160 x 2 = 480.

Survey of economic activity **Base-line survey for the year yyyy**

Company: _____< *pre-filled by Statistical Office* >_____ Code: < *pre-filled* >

Address: _____

_____ Town: _____

Mail Address: P.O. Box _____ Town: _____

Contact information of the department responsible for monthly provision of the information

Name of contact person: _____ E-Mail _____

Contact telephone: _____ Fax: _____

Value of sales from own production in yyyy: {national currency} _____

The value of sales in yyyy forms the basis for the computation of weights in the index system.

The production quantities are then used to calculate (aggregate) changes in industrial output from month to month and year to year.

Product line: _____
(if more than one product line, use separate form for each)

Production
Item description: _____
(provide description of the most common item or item-group in this line produced during the year)

Unit of measurement: _____
(provide your standard unit of measurement)

Quantity produced during the year yyyy: {yyyy+1}:

Quantity produced during the month of:

January yyyy+2	July yyyy+2 . .. .
February yyyy+2	August yyyy+2 .. .. ..
March yyyy+2	September yyyy+2 . ..
April yyyy+2	October yyyy+2 .. ..
May yyyy+2	November yyyy+2 .. ..
June yyyy+2	December yyyy+2 .. ..

Prices
In order to compile an index series for producer prices, it is necessary to obtain month-end prices for the same product/packaging from period to period. For each product line, the most common product/size should be quoted.

Please indicate below the appropriate description:

Pricing item _____

Unit of measurement _____

Price at last day of the year yyyy {currency} _____   {yyyy+1} {currency} _____
(invoiced price involving most transactions)

Price at last day of the month:

January {yyyy+2}	July {yyyy+2} . .. .
February {yyyy+2}	August {yyyy+2} .. .. ..
March {yyyy+2}	September {yyyy+2} . ..
April {yyyy+2}	October {yyyy+2} .. ..
May {yyyy+2}	November {yyyy+2} .. ..
June {yyyy+2}	December {yyyy+2} .. ..

General information
Number of employees at the end of the year _____
Number of days worked during the year _____
Number of shifts worked: 1 / 2 /3 per day or _____ during the year

Please return this form to: Or fax it to: ....
Statistical Office   Or e-mail to: ....
P.O.Box xxxx
Town   *For information, call us at ....*

Annex 3: Instructions for completing the monthly survey questionnaire

On the basis of the information obtained through the baseline survey, one, two, or three questionnaires will be provided each month for reporting the key information for the selected product line(s). All standard information is pre-filled in by the Statistics Office, as shown in the example provided below for the company, Coca Cola.

Under normal circumstances, only six items of information (show on the example form below) are sought for each of the selected product lines, namely:

1) The quantity of production of the selected item

2) The total value of production of all items of the product line

3) The price of the selected pricing item at the end of the month

4) The number of employees engaged on the selected production line

5) The number of days worked during the month

6) The number of shifts worked. This can be indicated as one, two or three shifts per day when it has been standard throughout the period. Otherwise, the total number of shifts worked during the year should be reported. For instance, if there were 320 working days and operations during 160 days were on a single shift basis, while the remainder was on double shifts, the total number of shifts worked is 160 x 1 + 160 x 2 = 480.

As there tend to be changes in product lines, production patterns and most common packaging and sizes over time, the monthly questionnaire also needs to keep track of these, where practicable.

The most common change is with respect to pricing items, as packaging styles, materials and sizes tend to change over time. For example, in the case of Coca Cola, the current standard 300ml glass bottles may be replaced by 500ml plastic bottles in a few years time. In such a case, it is important to modify the "quality" of the pricing item accordingly. For this purpose, the questions on the new pricing items and the start of new production lines are repeated every month.

In case another item has become the most important one, this should become the pricing item. In order to chain the index, the price of the new item at the end of the previous month is requested in addition to the price at the end of the current month. The description of the new pricing item should be such that it uniquely identifies a particular packaging of the selected product.

Finally, if large fluctuations in production have taken place during the month (for instance, the regular shut-down of an oil refinery for maintenance) this should be reported in the space provided.

After all the information is completed, the form should be returned to the Statistics Office by mail, fax, e-mail or any other appropriate means of communication. This should preferably be done within one week of the end of the month.

Monthly survey of economic activity   for the month of _____ 200_

Company: _____   Code: _____

Product line: _____

Production   Product Code: _____
Item description: _____
Unit of measurement: _____
Quantity produced during the month _____
Total value of production during the month {currency}_____

Prices   Pricing Item Code _____

Pricing item _____

Unit of measurement _____

Price at on the last day of the month: {currency}   *(invoiced price involving most transactions)*
_____

Is the pricing item still the most important product? If not, describe new pricing item below:

New pricing item _____

Unit of measurement _____

Price at on the **last** day of this month {currency} _____

Price at on the last day of **previous** month {currency} _____

General information
Number of employees at the end of the month _____
Number of days worked during the month _____
Number of shifts worked: 1 / 2 /3 per day or _____ during the month
Were new product line(s) started during the month? Yes / No

Please indicate below any special reasons for large fluctuations in production:
_____
_____

Please return this form to: Or fax it to: ....
Statisticsal Office   Or e-mail to: …..…..…..…..…..…..….
P.O.Box xxxx
Town   *For information, call us at . … … …*

Annex 4:   Introduction to the new system of production and producer price indices

The following draft text provides an example which could serve as the basis of an information note to be circulated by a Statistics Office to establishments requested to assist in providing the data needed for a new system of production and producer price indices:

Statistics Office of Country X

Introduction to the new system of production and producer price indices

"The Statistics Office is charged with providing accurate and up-to-date statistical information on major aspects of society. This includes the preparation of a range of economic statistics to monitor the economic developments in the country.

In order to improve the statistical information, the Office is currently undertaking a complete revamp of its economic statistics. This includes the rebasing and expansion of the national accounts as well as carrying out a new industrial census.

So far, the National Accounts compilation is centred on annual data sets, which are obtained largely from annual reports of the corporate sector, budgets and accounts of the government sector, and large-scale surveys of household sector activities. By its nature, such information becomes available only some time after the end of the reference period.

For effective management of the economy, as well as policy formulation, it is necessary to have information available as early as possible, with initial estimates preferably available by the end of year the under review. Furthermore, it is becoming regular practice to compile quarterly national accounts. For both of these activities, the use of "short-term indicators" for the measurement of changes in levels of economic activities is necessary.

Short-term indicators come in many guises and their commonality lies in the assumption that a particular time series gives a good indication of the change in the level of activity (volume or value of outputs) and/or the prices in the activity it represents. In general, the assumption of representativeness of an indicator should be checked from time to time as, in many cases, relationships between indicators and activities change over time.

The regular measurement of changes in various industrial activities is accomplished most easily with the help of a system of producer price and production indices. The producer price

indices provide information for the conversion between constant and current prices. The production indices measure the changes in volumes of production.

The strength of an index system lies in its ability to measure changes with a minimum of current information collected. The data that are collected are linked to the prevailing structure through its weights system. Under normal circumstances, this provides accurate estimates of developments in the short to medium-term as production structures generally change over time. It is generally recommended that re-basing should take place at five-yearly intervals. "

For countries with a VAT system, the following paragraph may be added:
"However, in our country we may consider ourselves fortunate in that we can use the annual aggregates on output and value added by industry generated by the VAT service from its database. This allows the Statistics Office to update the weights system for the index series on a continuous basis so that chained indices can be compiled. These types of indices capture both the change in production levels, as well as changes in industrial structure."

Annex 5:   Description of the PRIMA software

PRIMA[5] is a software system developed for the processing of various types of indices. The approach used is extremely flexible. It has a modular design and a generic structure. It is largely platform-independent and it has been implemented on a range of database systems, including MySQL and Oracle. Its main characteristics are described below.

A5.1   Modular design

The system consists of a number of separate modules:

- Data entry module (PDA and/or desktop)

- A separate, application specific, module for each application and / or country

- Indexer (desktop and/or web-based)

- Analytical module (entirely web-based)

A5.2   Generic structure

The system is data-independent; it takes the structure for a particular application from the structure of its (hierarchical) weights files.

The system is flexible:

- It allows for an unlimited number of quotations

- It allows for an unlimited number of items, product lines, enterprises, regions and so on

- It allows for weekly, monthly, or a less frequent cycle of observations

- It has eight levels of aggregation from the Item-level upwards

- It allows for separate weights files for each period

---

[5] PRice (or Production) Index MAnagement System

- It generates Laspeyres, Paasche or chained indices as required

- It accommodates up to eight different weights sets within a single weights file so that indices for different groups (e.g. export and domestic items; different income groups in a CPI) can be calculated separately

A5.3   Three modules of PRIMA

Data entry module

The main features of the data entry module are:

- Data entry can be done in various ways, namely on PDAs in the field, on the desktop in the office and, if the system is set up for it, online directly into the database.

- The desktop data entry screen is an exact representation of the product lines, items, and their qualities enumerated within each enterprise in a particular period. These features are also implemented on the PDA but, due to the limited screen size, the lay-out is different.

- The production (and prices) for the previous period are displayed along with the product/ quality information for ease of reference.

- The supervisor can set an acceptable production and price range. All production volumes and prices outside that range will be marked. The data entry operator (or the enumerator using a PDA) can only proceed after explicitly accepting the new price (marking a separate field). The transaction will then be recorded in a separate list of suspected movements for review and editing by the supervisor.

- The changes in quality for all items are recorded in a separate module. The new qualities will be used in later questionnaires and data entry schedules so that the information available is always up -to-date.

- The data entry module is directly linked to the new questionnaire. The latter can be generated (in the form of a PDF file) and printed or created as an XML file for data entry using PDAs (hand-held computers).

- It includes only those enterprises, product lines and items for which production (and prices) information have been collected during the previous period and the changes in qualities recorded for that period.

After data entry and editing is completed, the data are ready for index calculations.

Indexing module

The index calculations are fully automatic. After starting the Indexer, a selection screen comes up. The appropriate selections should be made and the indices will then be calculated. The following items need to be selected:

- **Survey set**: The indexer is generic. Different data sets may be present (Consumer Price Index (CPI), Producers price index (PPI), Wholesale Price Index (WPI), Index of Industrial Production (IIP), etc.).

  Note: If only one system is loaded, it is automatically selected.

- **Weights set**: Multiple weights sets may be present for a system. For instance, a PPI may have weights based on both output and value added.

  Note: If only one weights set is available for a survey set, it is selected by default.

- **Index type**: A choice between Laspeyres, Chained and Paasche types is available.

  Note: Paasche indices will be calculated using the (base period) weights for the current period or the latest available period. If only a single base period is available for a survey set, the Paasche index degenerates to Laspeyres by default.

- **Number of weights columns**: By default one; needs only to be set if more than one weight is available for calculations (e.g., low, medium and high income groups).

After the selections have been made, the base period and the periods for which the indices need to be calculated should be selected.

- **Base period**: First, the base period needs to be selected from the list of options available. This list is automatically updated when new weights sets are added or old ones removed. If only a single base period is available, that period is selected automatically.

  Note: The base period is defined in terms of production or pricing reference. For instance, 1997-09 as a base period means that the production or prices of September 1997 are used as reference in the calculations. Similarly, 2002-0 means that the average production or prices for the year 2002 are used as reference.

- **Periods**: All periods with data available for the weights set selected are shown in this data block. One or more periods (with a maximum of 24 per run) can be selected for index calculations. If the appropriate box has been ticked, the results are also stored in an archive. New data will be added and recalculated data will replace those already on the archive file.

- **Calculate indices**: After all selections have been completed, this option should be pressed. The progress of the calculations is shown at the top of the screen.

- **View / Export**: The results of the calculations can be viewed by pressing this option. If pressed without calculations, it will show the latest data in memory. Data can be viewed at different levels of aggregation (from full detail at item-level upward to overall total only). The export option transfers the data with the detail shown in the view to a DBF (dBase) file for use in any database, word processor or spreadsheet.

Analytical module

The analytical module is web-based and therefore can be run from any location, as long as the detailed data in the archive of the indexer have been put into the database. It allows the preparation of a wide variety of reports. The data can also be used for forecasting and regression analysis, for which a number of facilities have been built in. In addition, the selected data can be exported in the appropriate format for direct use in the extensive Gretl analytical package[6].

The following analysis and tabulation tasks are possible:

- Calculate contributions by products and groups to changes in the index from month to month

- View indices by period for any number of periods and any level of aggregation

- X-11 seasonal adjustment with ARIMA (Auto-Regressive Integrated Moving Average); Forecasting and backcasting[7]

- Multiple regression for correlating exogenous time series data with selected index branches and forecasting using the TRAMO algorithm year-to-year and month-to-month series

- View as HTML, text or graph

- By filtering, aggregate product indices in any way required for the analysis, for example, identifying all activities that use a particular packing material to estimate the expected demand for such materials

The interface of the analytical module is step-by-step and selection options are made available depending on choices made so far.

---

[6] This package is open source and therefore free. Its functionality is comparable to that of SPSS, SAS and Stata. For a comparison of the packages, see http://en.wikipedia.org/wiki/Comparison_of_statistical_packages

[7] Using X-11 and TRAMO algorithms implemented by the Research & Development Cell of the National Bank of Belgium.

In the first selection, one or two data sets provide an option to link two data sets for various types of analysis. This option is only available when compatible datasets are selected.

In the second step, the period(s) need to be selected, along with the level of detail and the number of sub-indices. As in the Indexer, the list of periods and levels available depends on the data contained in the data set that was selected.

At this point, filtering to select only partial data from the set is also started. Two options are available, namely, picking the indices at higher levels from the database, or aggregating them anew using the relative weights for the selected items or groups within the data set only. The filtering selections go through all levels of the indices, from the top down.

After the selection has been completed, the various calculation options listed above can be selected. After review of the selection results, the data can be saved either as comma-delimited text files, in html-format or as JPEG graphs. These can be transferred for use in any database, word processor or spreadsheet package.

The TRAMO multiple regression and forecasting option requires the availability of some exogenous time series, which can be locally entered or selected from a list of available series (including, for example, crude oil prices and exchange rates). The results of this analysis, which uses pre-set or optimised values for most parameters, can be saved in a file that can be used directly in the Gretl analysis package where all these parameters can be changed for sensitivity analysis and other purposes.

Chapter 6: Statistical Indicators of Industrial Performance

12.0   Introduction

The results of a regular Industrial Survey, using data items described in Chapter 2, provide a wide range of statistics that can be used for the compilation of various statistical indicators for the analysis of industrial performance. The International Recommendations for Industrial Statistics 2008 (IRIS-2008) include a chapter on performance indicators encouraging NSOs to compile a set of statistical indicators in order to meet the increasing demand from data users in the business community and from policymakers. Performance indicators primarily measure the economic efficiency of industrial production with respect to the growth of output in relative terms of input. For the business community, performance indicators provide information needed to assess the current economic situation and to indicate the prospective direction in comparison with international economic trends. Performance indicators also help to assess correctly the economic situation of a country as a basis for designing and implementing an effective fiscal or monetary policy. An arbitrary decision or tardy diagnosis of business trends may result in unsound decisions with serious economic consequences.

Performance can be evaluated at the firm level, at the ISIC branch level, as well as at the level of the industrial sector in relation to other economic activities or in relation to the economy as a whole. Statistical indicators also allow for the evaluation of industrial performance across countries, thus facilitating comparative analysis. In terms of the computation method, IRIS 2008 has broadly defined three types of performance indicators, namely; (a) growth rates, (b) ratio indicators, and (c) share indicators. Construction of such indicators can be based on a wide range of variables for different analytical purposes. In fact, the compilation of performance indicators is that part of data analysis which is the final stage of any statistical operation.

Analysis of industrial performance is also carried out using complex multivariate statistical analysis and econometric models. The purpose of these Guidelines, however, is to provide staff in National Statistics Offices and line ministries with the basic data analysis tools that can be applied to data collected from a regular survey programme of official statistics.

13.0 Data sources

The number and type of indicators that can be developed for performance measurement are determined by the data availability in the national industrial statistics database. Based on the International Recommendations on Industrial Statistics and established statistical practice in many countries, it is reasonable to assume that data for the following variables are readily available in the national industrial statistical database or in the survey results by ISIC at 3 or 4 digit level:

1) Number of employees

2) Compensation of employees

3) Gross output

4) Intermediate consumption

5) Value added

6) Gross fixed assets at the end of the reference year and the gross fixed capital formation

7) Index of industrial production

The above indicators provide information about domestic production. For a comparative analysis of production with domestic consumption, it is necessary to relate the production data to foreign trade statistics on the following items:

8) Export of manufactured goods

9) Import of manufactured goods

Given the coverage of export-import data, it should be possible to derive the foreign trade balance at national level with a fair degree of reliability. For an overall assessment of the performance of the manufacturing sector in relation to the economy as a whole, data are also needed for the following indicators, which are also generally available in national databases:

10) Population

11) Gross domestic product (GDP)

12) Manufacturing value added (MVA)

All of these indicators are familiar to NSOs and are clearly defined in different international recommendations. Most of the NSOs regularly compile and publish these data. The performance indicators described in the following paragraphs are based on the above variables. The analysis of industrial performance involves one further step in order to derive new indicators from the data that are already available.

14.0 Types of performance indicators

Industrial performance is an outcome of various social, economic and technological factors. These Guidelines cover the three most important dimensions of industrial performance, namely productivity, structural change, and competitiveness. Each of these dimensions is represented by a number of indicators described below.

14.1 Productivity

Productivity is an important measurement of the efficiency of production derived from the ratio of the real output to input over a period of time. A precise measurement of productivity is not so straightforward, especially due to the complexity of input components. In economic statistics, inputs are disaggregated into two major groups, namely primary inputs (also known as factors of production) comprising labour input and capital input; and intermediate inputs (materials and supplies, energy, industrial services, etc.). The change in productivity occurs as a result of the change in the efficiency of the use of all these inputs. A modest estimation of the combined effect of a change in primary inputs on output is obtained from econometric models of total factor productivity. In this chapter, we mainly present indicators of single factor productivity and describe a very simple approach to constructing a multifactor productivity index. Construction of these indicators is based on the relation of different output and input components at the national, industrial sector, and enterprise levels.

14.1.1   Manufacturing value added per capita

The most usual measure of productivity is the ratio of output to the entire population of a country. In this case, the population is taken as a producer and output as the result of its productive activities. The manufacturing value added (MVA) per capita as a relative value of net manufacturing output to the population size indicates the level of industrial development achieved by a country. MVA is estimated as the total of value added of all manufacturing activities in the country. Unlike gross output, MVA is free of double counting as it excludes the cost of intermediate consumption. At the same time, it is a comprehensive measure of production as it includes the cost of labour, consumption of capital

(depreciation) and the operating surplus. MVA is measured at basic prices so it is free of the variation caused by the ever changing rate of commodity taxes.

MVA per capita is computed as:

$$\text{MVA per capita of a country} = \frac{MVA}{Population} \qquad (1)$$

MVA per capita as a component of GDP per capita can also be estimated in direct relation to GDP. MVA per capita as a derivative of GDP per capita and MVA share in GDP is given by:

$$\text{MVA per capita} = \text{GDP per capita} \times \text{MVA share in GDP} =$$

$$= \frac{GDP}{Population} \times \frac{MVA}{GDP} = \frac{MVA}{Population} \qquad (2)$$

The above relation synthesizes the overall role of the manufacturing sector in the economy and its contribution to the income of the population.

Table 1: Growth of MVA per capita at constant 2000 US$ by country groups

	1990	1995	2000	2005	2008
Industrialized Countries	**100**	**103.2**	**120.4**	**127.1**	**123.4**
CIS	100.0	46.9	51.4	70.9	90.3
European Union	100.0	100.8	114.8	117.9	121.0
North America	100.0	113.1	142.1	145.9	120.8
East Asia	100.0	102.3	110.1	120.0	126.5
Others	100.0	96.8	102.3	101.9	112.3
Developing countries	**100.0**	**125.9**	**147.7**	**190.5**	**216.5**
Asia excluding China	100.0	118.3	135.2	165.7	154.1
China	100.0	198.0	302.3	490.5	686.8
Africa	100.0	95.9	111.0	116.3	124.4
Sub-Saharan Africa	100.0	88.1	93.1	99.0	110.6
Latin America	100.0	111.8	110.5	122.0	126.0
Least-developed countries	**100.0**	**95.1**	**114.9**	**143.7**	**175.0**
World	100.0	103.1	116.2	125.0	124.9

Source: UNIDO MVA database

14.1.2 Valued added per employee

Value added per employee is a more concrete measure of the relation of output to labour input. It indicates the average amount of value added produced by an employee. This is the simplest approximation of the rate of labour productivity in manufacturing:

$$\text{Value added per employee} = \frac{\textit{Value added}}{\textit{Number of employees}} \qquad (3)$$

A more comprehensive measure of the total size of employment in industries is the number of persons engaged, which, in addition to the number of employees, also includes the

number of working proprietors and unpaid family workers. It is quite common to compute the average value added per person engaged for consistency checks of the survey results. However, as a measure of labour productivity value added per employee has advantages. First, benefits received by the working proprietors and unpaid family workers from working in the industry do not necessarily depend on their labour input to production. Second, the average number of hours worked in a certain period by the working proprietors and unpaid family workers may significantly deviate from that of employees.

The growth of value added over time depends on the change in the quantity and value of goods and services produced. For a measure of the real change in productivity, it is convenient to calculate the index of labour productivity as a ratio of the production index, which shows the growth in the volume of production to the index of the number of employees:

Index of labour productivity = $I_{LP} = \dfrac{I_Q}{I_L} \times 100$ (4)

where:

I_Q and I_L = index of industrial production and the number of employees respectively for given period

Data for production indices and the number of employees may also be available from quarterly surveys, in which case this index can be included in the set of short term indicators.

14.1.3   Valued added per hours worked

Among the indicators based on labour input, value added per hour worked provides the most precise estimate of labour productivity, although data on hours worked may not always be available. This ratio is obtained by relating the amount of value added for a reference period to the total number of hours worked during the same period:

Value added per hour worked = $\dfrac{Value\ added}{Number\ of\ hours\ worked}$ (5)

The main advantage of this ratio is that it is free from the effects of the variation in hours worked by individual employees. A statistical unit may have part time, seasonal and full time workers. Some employees may work overtime while others may go absent. When the value added is related to the number of employees, these differences are not taken into account. NSOs are generally encouraged to collect the data on the number of hours worked in order to measure labour productivity more precisely.

14.1.4 Value added per unit of capital

Capital here refers to the gross value of fixed assets at the end of the reference period as defined in IRIS-2008. The value added per unit of capital is calculated as:

$$\text{Value added per unit of capital} = \frac{\textit{Value added}}{\textit{Gross fixed assets}} \qquad (6)$$

This ratio measures the productivity of capital in the reference period. The result of capital replacement in productivity can be obtained by the ratio of the annual increase of value added to the gross fixed capital formation, which is given by:

$$\text{VA/capital increment ratio} = \frac{VA_1 - VA_0}{FA_1 - FA_0} \qquad (7)$$

where:

VA_0 and VA_1 = value added for year 0 and 1

FA_0 and FA_1 = gross fixed assets at the beginning and end of year 1

The limitation of this indicator is that, in the event of negative additions to gross fixed assets, especially at the sub-sector level, it would be difficult to interpret the ratio. Therefore, the preferred method is to construct an index similar to the *index of labour productivity*, which would appear as:

Index of capital productivity = $I_{KP} = \dfrac{I_Q}{I_K} \times 100$ (8)

where:

I_K = index of the value of gross fixed assets

14.1.5 Capital per employee

This indicator measures the relation of the factors of production, which is calculated as the ratio of the value of gross fixed assets to the number of employees:

$$\text{Capital per employee} = \frac{Gross\ fixed\ assets}{Number\ of\ employees} \qquad (9)$$

In comparison across sectors and sub-sectors, the capital per employee indicates the factor intensity – the relative importance of one versus the other. In statistical practice, there is no clear demarcation to divide industries in the dichotomy of labour-intensive or capital-intensive. A possible measure is to apply the median value. The following chart is drawn from the data obtained from the 2007 Industrial Census of Thailand.

Figure 1: Value of the fixed assets per employee in manufacturing sectors in Thailand, 2006
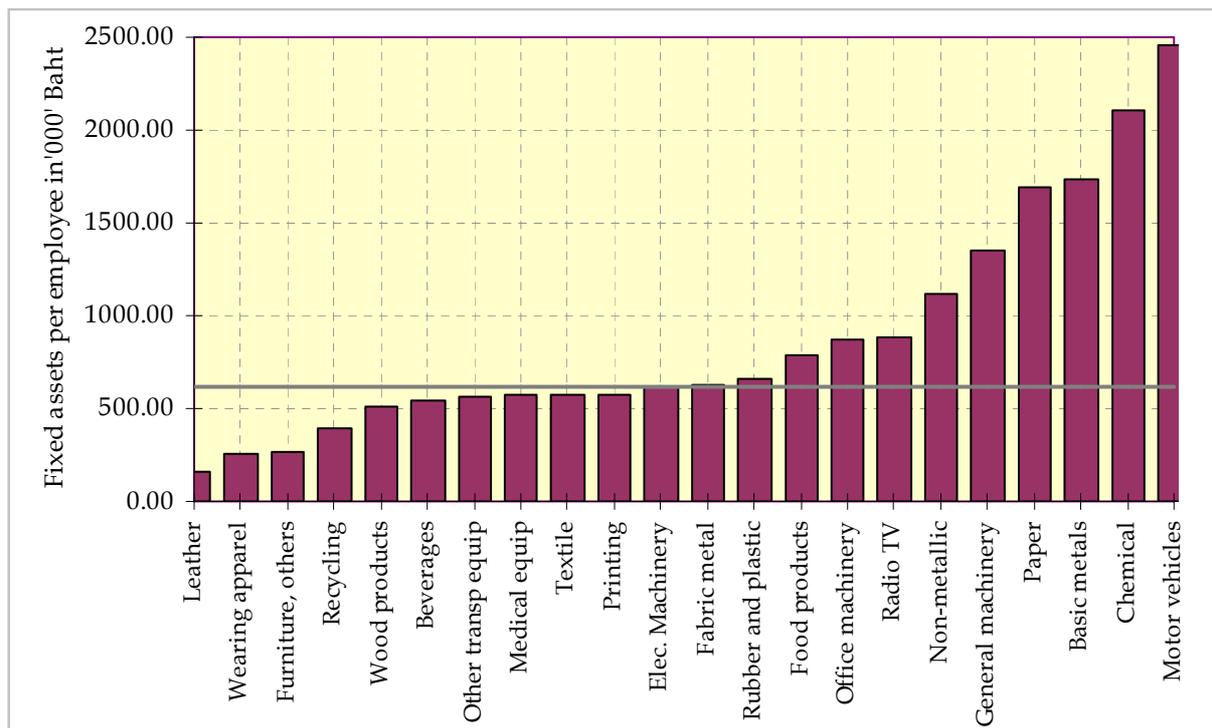
In the above chart, the manufacturing sectors with the value below and above the median line can be considered as labour-intensive and capital-intensive respectively relative to one another.

The relation of the factors of production changes over time and this can be measured by the *capital intensity index* computed as the ratio of two indices related to capital and labour.

$$\text{Capital intensity index} = I_{INT} = \frac{I_K}{I_L} \tag{10}$$

where:

I_L and I_K are defined above in relations (4) and (8)

This index shows the direction of the shift of one or another industry in terms of factor intensity. An increasing value in this index over time shows that an industry is becoming capital intensive.

14.1.6 Multifactor productivity index

This index refers to the ratio of the production index to the total of labour and capital indices adjusted by their proportional contribution to value added. The formula of the multifactor productivity I_{MF} index is given by:

$$I_{MF} = \frac{I_Q}{(I_L \times S_L) + (I_K \times S_K)} \tag{11}$$

where:

S_L = share of compensation of employees (as labour input) in value added

$S_K = 1 - S_L$ = share of other components (capital input) in value

I_L and I_K are defined in relation (4)

The figure for value added from the Industrial Survey data is normally derived as the balancing item of output and input. Therefore, precise figures for each component of the value added might not be available in the survey results. However, data on compensation of employees is reported separately (refer to Section 2, entitled Employment, wages and salaries, hours worked, in point 3 of Chapter 2). As the remaining components relate to capital input (consumption of fixed capital and operating surplus), it is valid to assume the share of capital input is the residual after deducting the labour cost from the value added (i.e. $S_K = 1 - S_L$).

For the multifactor productivity analysis, it would be convenient to present the output data from the Industrial Survey results in relative terms by components: material input, labour input, and capital input (as the residual). Such decomposition also enables cost analysis across manufacturing sectors and comparison of data with the output structure of other countries.

The following chart is based on the Industrial Survey results of selected countries in Latin America and the Caribbean. Data refer to 2005, except for Chile and Trinidad and Tobago, where 2006 data are presented.

Figure 2: Percentage share of input components in the readymade garment sector of selected countries in Latin America and the Caribbean
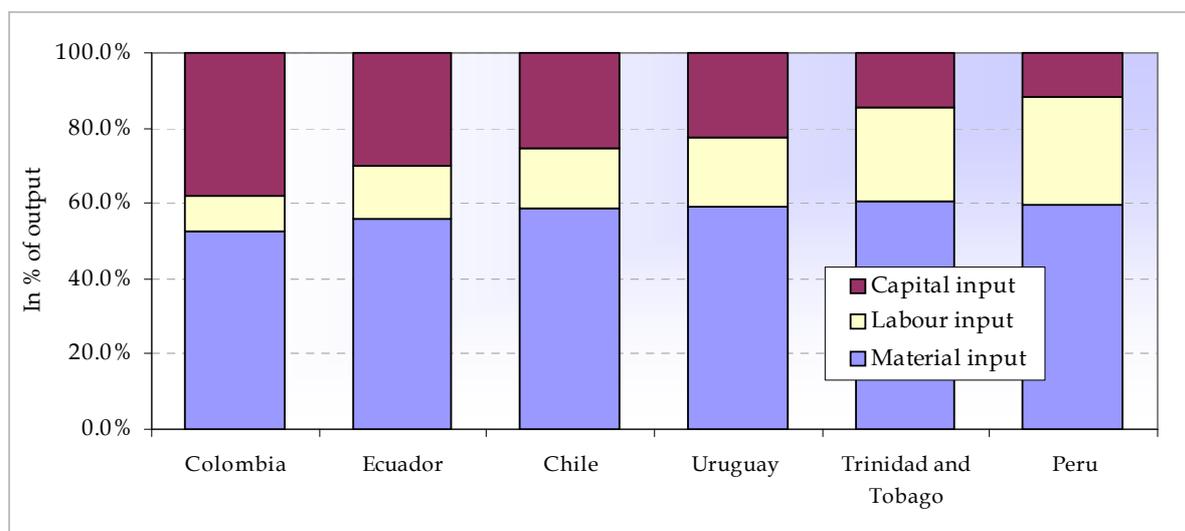
14.1.7   Value added output ratio

As also evident in the above graph, it is material input which accounts for the largest portion of gross output. While the indicators mentioned so far are intended to explain the role of production factors, the value added output ratio measures the efficiency of the use of material inputs. The main advantage of this indicator is that it can be calculated at any level – at the manufacturing level, at the level of sectors and sub-sectors, as well as at the enterprise or establishment level. The value added/output ratio is calculated simply as the relation of value added to gross output:

$$VA/output\ ratio = \frac{Value\ added}{Output} \qquad (12)$$

This measure is often presented as the percentage share of value added in total gross output. The ratio is derived from the basic equation applied in economic statistics, which describes value added as the measure of production net of input:

$$Value\ added = Output - Input \qquad (13)$$

The relevance of this ratio as a productivity measure is explained by its relation to the input-output ratio. The VA/output ratio is inversely proportional to the amount of input per unit of output; hence, it represents a more productive use of the input in the production process.

$$VA/output\ ratio = 1 - \frac{Input}{Gross\ output}$$

(14)

The VA/output ratio is a more stable and robust estimate in comparison with the absolute value of output and input. Thus, it is widely applied in order to check the data consistency in survey results. The value added/output ratio varies significantly according to the sector. Generally, it is lower for resource-based sectors, such as agro-industries and basic metals, but higher in technologically-advanced sectors, such as the pharmaceutical industry and machinery production.

14.2   Structural change

Structural change refers to the dynamism of the cross-sector relationships of the principal indicators and is measured as the shift of sector shares over a considerable period of time. Structural change is closely connected with industrial growth. On the one hand, structural change may accelerate growth while, on the other hand, any growth may result in significant structural change. Therefore, both share and growth indicators are presented, as well as an indicator of regional structure and the position of manufacturing in the total economy.

14.2.1   Change in sector share

At the most elementary level, structural change is the shift in the share of individual sectors over a certain span of time and can be measured by the difference in the share of a sector in the total value of the variable being analysed, as shown below;

$$d_s = s_{i1} - s_{i0}$$

(15)

where:

d_s = the difference of the share between two periods, 0 and 1

The share of a sector in total manufacturing is calculated as:

$$s_k = \frac{x_k}{\sum\limits_{i=1}^{n} x_i}$$ (16)

where:

s_k = share of k-th sector in total

x_i = value of the variable being analysed (employment, output, capital etc.) for i-th ISIC branch

n = number of ISIC branches in the observation

The percentage share is derived as:

$$s_k = \frac{x_k}{\sum\limits_{i=1}^{n} x_i} \times 100; \, k = \overline{1, m}.$$ (17)

The sector shares are calculated at the different levels required for the purpose of data analysis. If the share is calculated for a group of sectors instead of individual sectors, the numerator of the above relation is replaced by the group sum. Comparison of the share of sectors or group of sectors over time indicates the direction and intensity of the structural change in manufacturing. In order to obtain meaningful results from the structural change analysis, UNIDO has developed a number of classifications based on grouping several ISIC branches together, such as:

1) Resource-based industries

2) Agro-industries

3) Production of ICT goods

4) Medium and medium-high technology sectors (MHT)

Details of these sectors, together with their corresponding ISIC codes, are given in Annex 1.

In the early stage of industrialisation, the structure of manufacturing is such that the processing of raw materials produced by the primary sectors plays a dominant role. At the advanced stage, production becomes technology-intensive and the share of resource-based sectors in total production gradually declines. The example illustrated below refers to the Republic of Korea, a country that has achieved great success in the last 40 years through its transformation into a modern industrialised economy.

Figure 3: Percentage share of selected ISIC branches in total value added in the Republic of Korea
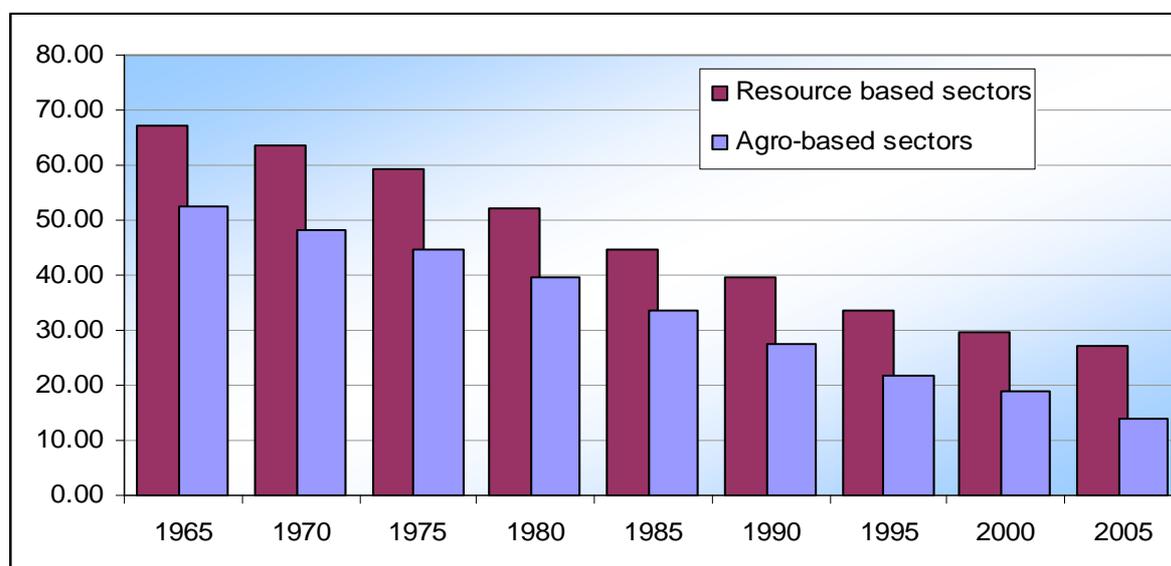
Source: UNIDO database - INDSTAT2

Figure 3 shows that the share of resource-based sectors, which process the primary products from agriculture and mining and quarrying, gradually declined from almost 70 per cent of total net output in 1965 to less than 30 per cent in 2005. In a similar way, the agro-industrial sector's share declined from 53 per cent to only 14 per cent. Consequently, the secondary processing and high-technology sectors accounted for more than 80 per cent of the country's total net output in 2005.

The change in sector share described here refers to an individual ISIC branch or a group of ISIC branches. For an overall assessment of the structural change in manufacturing, we can measure the combined effect using various indicators described below.

14.2.2   Coefficient of absolute structural change

This coefficient measures the structural change in industry by comparing the total difference of the change in shares for all sectors with respect to a variable of interest, such as employment, capital or output over two periods. It is an average value of absolute changes across manufacturing branches. The coefficient of absolute structural change for a variable *x* in a given time is calculated as:

$$d(x)_{abs} = \frac{\sum_{i=1}^{n} \left| s^{x}_{i1} - s^{x}_{i0} \right|}{n} \tag{18}$$

where:

s^{x}_{i} = share of the variable *x* for the *i*-th sector; suffix 0, 1 stands for period 0 and 1, respectively

n = number of breakdowns (sectors in observation)

The coefficient equals zero if there is no change in any of the branches over the observed period. This coefficient may take a value up to $2/n$ at the maximum. Obviously, a lower value closer to zero indicates the lack of structural change and conversely, a higher value indicates more significant change in the structure of industrial sectors.

14.2.3   Coefficient of relative structural change

Structural change can also be measured in relative terms. For this purpose, the absolute difference is replaced by the ratio of selected variables for the later year to that of the earlier year. The above relation can be restructured to measure the relative change as given by:

$$d_{rel} = \sum_{i} \left| \frac{s_{i1}}{s_{i0}} s_{i0} - s_{i0} \right| = \sum_{i} \left(\left| \frac{s_{i1}}{s_{i0}} - 1 \right| \times s_{i0} \right) \tag{19}$$

Instead of taking the absolute value to avoid the zero-sum of the structural change, the coefficient can be based on the squared differences.

$$d^2{}_{rel} = \sum \left(\frac{s_{i1} - s_{i0}}{s_{i0}} \right)^2 \qquad (20)$$

This coefficient shows the average deviation of the sector shares in the later year from that of the earlier year. Again, the value of squared difference will equal zero when there is no change across all the sectors. Any significant change in structure would result in a higher value of the coefficient (from zero). However, this coefficient has no upper limit; thus it cannot be defined in a range.

14.2.4 Integral coefficient of structural change

This index measures the combined effect of the sector shifts on the overall structural change in industry. The formula of the coefficient is given by:

$$d_{int} = \sqrt{\frac{1}{n} \sum \frac{(s_{i1} - s_{i0})^2}{(s_{i1} + s_{i0})^2}} \qquad (21)$$

The advantage of this index over those mentioned above is its interpretability. The value of the coefficient ranges from 0 to 1. The magnitude of the coefficient represents the scale of the structural change. The value of $d_{int} \leq 0.1$ would mean an identical structure in both time periods, $d_{int} \geq 0.9$ would indicate a complete reversal of sector distribution. For more realistic scenarios, a significant structural change would result in a value of $d_{int} \geq 0.5$.

14.2.5 Rank correlation

The rank correlation is a widely used method in the assessment of structural change, which shows the strength of the relation between the sector shares in two periods. The rank correlation coefficient of Spearman is given by:

$$\rho_s = 1 - \frac{6 \sum d^2}{n(n^2 - 1)} \qquad (22)$$

where:

d = the difference of ranks of manufacturing sectors (ISIC branches) in observed time periods by a chosen variable

n = number of sectors in observation

The procedure for the computation of the rank correlation coefficient is simple. First, manufacturing branches are ranked from 1 to n by a chosen variable, such as value added in the base year. Then, they are ranked again by the value of the same variable for the current year. Several sectors may have different ranks in the current year from those in the base year. The value of d^2 is computed as the absolute difference in the ranks in two periods, that is, $d^2 = (rank\_base - rank\_current)^2$. Further calculation is guided by the formula. The interpretation of the coefficient is such that when sector shares do not differ much between the two periods, they show a high degree of correlation and the value of ρ is close to 1. The value of ρ tends to zero for less correlated data, consequently sector shares are more diverse and there is significant structural change.

Figure 4 shows the change in ranks of manufacturing branches by industry value added in 1995 and 2005 for Brazil and Chile. It is clear that sector shifts were more significant in Brazil than in Chile. The figure shows the extent of the change in several ISIC branches based on the value of d^2.

Figure 4: Absolute change in the rank of ISIC branches by net output in Brazil and Chile
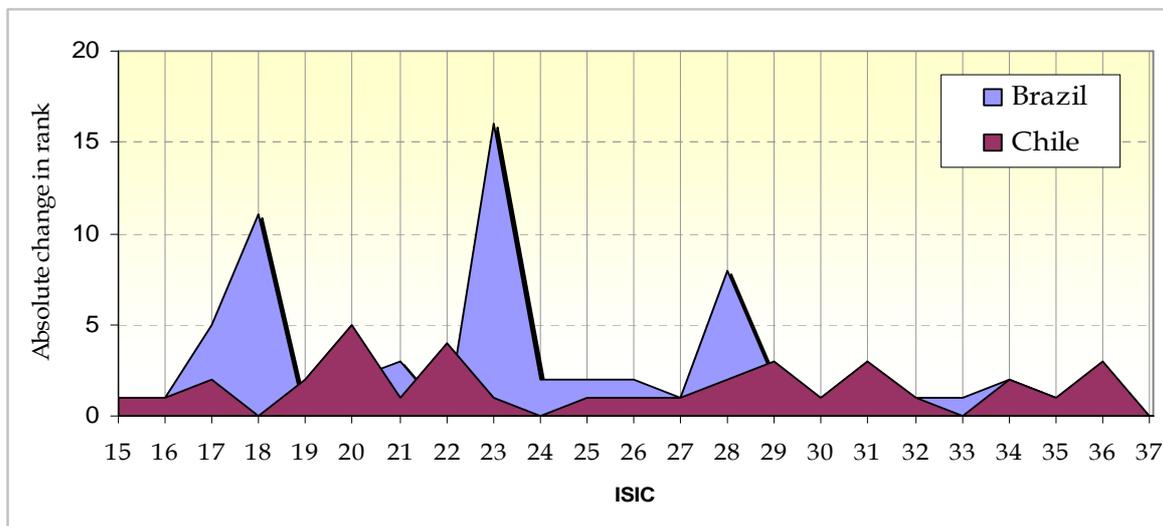
Source: UNIDO Database - INDSTAT2

The coefficient of rank correlation for Chile was calculated to be 0.95, which indicates highly correlated sector shares and marginal structural change between the two periods. Conversely, the coefficient was 0.73 for Brazil, indicating a less correlated and more diverse nature of sector shift in the observed ten year period.

14.2.6 Coefficient of diversification

Structural change and subsequent industrial growth lead to the diversification of production across a wide range of manufacturing activities. The least developed economies show a relatively narrow range of production specialised in a few products, often dictated by demand in the external market. As the economy grows and the domestic market expands, production diversifies significantly. The coefficient of diversification shows the extent to which the production is spread across different manufacturing branches and is based on the share of manufacturing branches in total output. It is calculated as:

$$C_d = \frac{\sum s_i \ln s_i}{- \ln n} \tag{23}$$

C_d equals zero when the value is concentrated in one branch of industry, which would mean a complete lack of diversification. The coefficient of diversification equals 1 when all the branches have an equal value, indicating perfect diversification.

In order to compare the dynamism of diversification, it is necessary to compute this coefficient over relatively longer intervals since its value may change only marginally over short periods, thus limiting its importance for economic analysis.

14.2.7 Regional disparity index

This index measures the extent of disparity in industrial development among the regions of a country. The index can be constructed in terms of any major indicators referring to resources utilized or results achieved. It is based on the deviation of actual value from the equal distribution of variables selected. The index is calculated as:

$$I_z = \frac{n}{2(n-1)} \sum_{i=1}^{n} /z_i/ \quad \text{and} \quad z_i = s_i - \frac{1}{n} \tag{24}$$

where:

I_z = index of regional disparity for z measure

n = number of regions in observation

s_i = share of the *i-th* region in total value of selected variable

The index has the advantage of simplicity for interpretation. When the distribution is fully equal over all regions, the value of z for each region is zero. Subsequently the index is 0, which means there is no disparity. The index is equal to 1 when one region has all the value and all other regions have 0, which would mean a complete disparity. While the index provides a consolidated measure of disparity, deviation of values for individual regions from the line of parity can be presented using a graphical illustration. The following graph depicts the employment distribution in the manufacturing industry of Ghana.

Figure 5: Distribution of the number of persons engaged in manufacturing by region
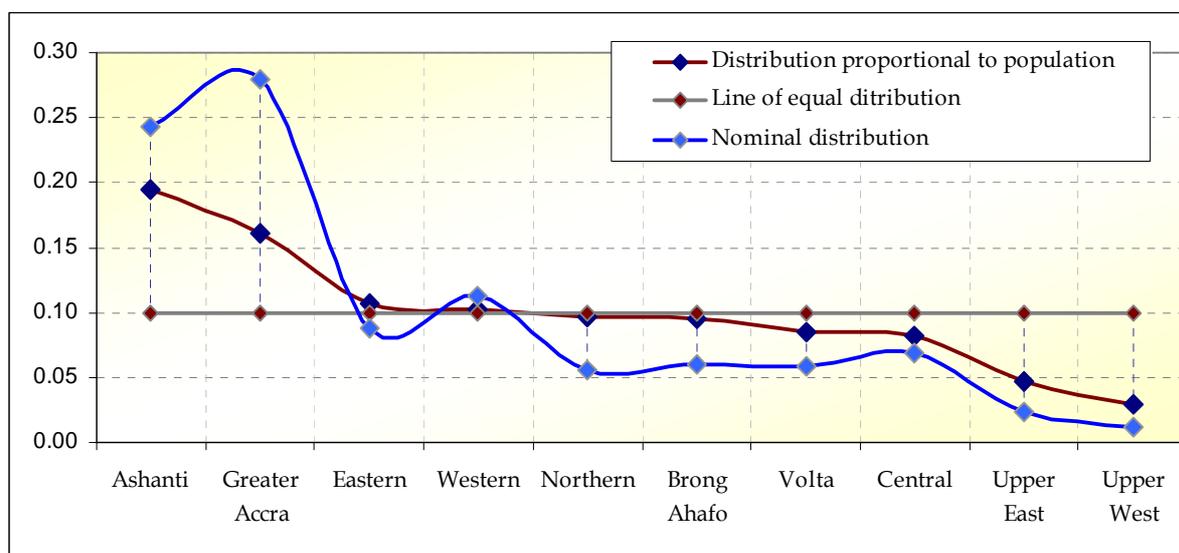
Source: GSS - National Industrial Census 2003, Ghana

In this graph, the nominal distribution is built on the share of regions in the total number of persons engaged in manufacturing sectors, while the proportional distribution is adjusted by the share of the region in total population of the country. In both distributions, the Ashanti and Greater Accra regions show a great divergence from the regions further inland.

14.2.8 Position of manufacturing in economy

The indicators described above provide statistics for the analysis of structural change within the manufacturing industry. The position of manufacturing in the economy and its contribution to the overall structural change in the economy can be measured by a number of share-related indicators. One of the most widely used is the percentage share of MVA in GDP.

$$\text{Share of MVA in GDP} = \frac{MVA}{GDP} \times 100 \qquad\qquad (25)$$

As a developing country industrialises, so does the contribution of manufacturing to the national economy and the share of MVA in GDP increase. In highly industrialised countries (see Figure 6), the opposite trend is observed for two main reasons: first, service sectors in industrialised countries tend to grow faster than manufacturing sectors and, second, there is a tendency to export the production process to developing countries through outsourcing in various forms.

Figure 6: Share of MVA in GDP in industrialised and developing countries
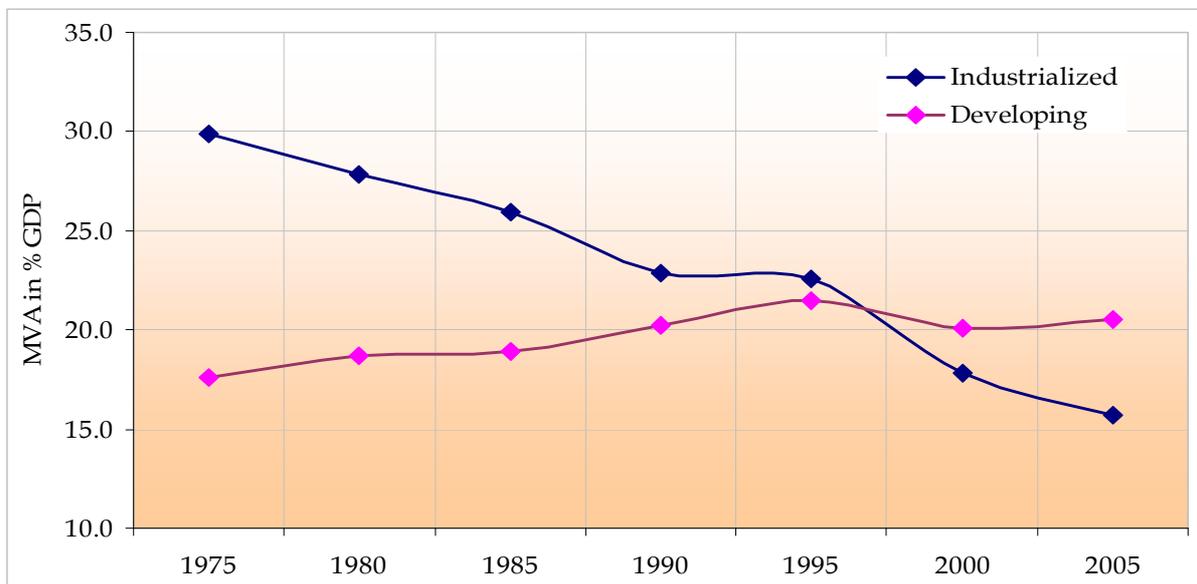
Source: UNIDO MVA Database

The relative expansion of manufacturing can also be measured in terms of other share indicators such as the number of persons engaged in manufacturing activities in relation to the total economically active population of a country, manufactured goods in relation to total merchandise exports and gross addition to fixed capital formation in manufacturing in relation to total gross fixed capital formation in all sectors. These indicators measure the contribution of the manufacturing sector to the overall economic development of the country. Subject to data availability, similar statistics could also be compiled for more specific areas such as the role of the manufacturing sector in research and development, in the production of ICT goods and services or in the consumption of energy resources.

14.3 Competitiveness

Statistical indicators of competitiveness measure the ability of a producer to sell its products in the market. Competitiveness is studied at the firm, industry or national level. Firm and industry level competitiveness can be analysed separately for the domestic and international markets. This chapter presents the indicators related to the performance of a country's industry in the external market.

Construction of these indicators is based on the assumption that the volume of exports is a good measure of a country's performance in the external market. As the volume of exports depends on domestic production (supply), minus domestic consumption (demand), the overall relation of domestic production to the external market can be based on the following equilibrium:

$$C = Y + M - X \qquad given \ that \ \ X \leq Y \ and \ M \geq 0$$
(26)

where:

C = derived value of apparent consumption

Y = domestic production

M = import

X = export

Theoretically, such a demand:supply balance can be built for each product. Similarly, the products can be aggregated to the level of industry branch. UNIDO maintains an international database on industrial demand-supply balance at the 4-digit level of ISIC (IDSB database). The data are obtained from the UNIDO INDSTAT4 database and the UN Comtrade database. Some of the indicators compiled by UNIDO for measuring competitiveness are based on statistics obtained from the IDSB database. These indicators are recommended for compilation at the national level as part of the performance indicators. The composite index of competitive industrial performance (CIP index), which has been regularly used in country rankings, is also described here.

Export ratio

The most generalised measure of competitiveness can be obtained from the export ratio as given by:

$$R_{X/Y} = \frac{X_i}{Y_i} \qquad (27)$$

where:

$R_{X/Y}$ = export ratio

X_i = export of goods of i-th product group or branch

Y_i = domestic production of i-th product group or branch

In the manufacturing sector, the export ratio represents the ratio of the export value of all manufactured goods to gross output, which is the sum of the value of goods and services produced by all manufacturing establishments resident within the boundaries of a country. A precise estimate of the export ratio can only be derived when both production and export data are collected from the enterprises. However, in most countries, production data are collected from Industrial Surveys based on figures reported by enterprises or their affiliated establishments while foreign trade data are obtained from statistics compiled by the customs office. A discrepancy arises because enterprise data on production includes services whereas custom data on trade does not. Nonetheless, the export ratio calculated from the two datasets is still a reliable approximation.

This ratio can be computed at the level of the manufacturing sector, sub-sector or product. The maximum value of $R_{X/Y}$ is 1, when a country exports all of its production of the i-th product group/industry. The ratio is zero when there is no export.

The analysis of the export ratio can be complemented with the ratio of domestic production to total consumption derived from the relation (26), which is called the self-sufficiency ratio:

$$R_{Y/C} = \frac{Y_i}{C_i} \qquad (28)$$

where:

$R_{Y/C}$ = ratio of self-sufficiency

Y_i = domestic production of i-th product group or branch

C_i = domestic consumption of i-th product group.

A country is considered to be self-sufficient and ready to export a product or products of an industrial branch, when the coefficient $R_{Y/C} \geq 1$. If the value of this coefficient is less than 1, it would imply that the country has to rely on imports to meet its internal demand.

14.3.1   Manufactured exports per capita

This is a relative measure of the export of manufactured goods of a country with respect to its population size. It is calculated as:

$$\text{Manufactured exports per capita} = \frac{Total\ manufactured\ \text{exports}}{Population} \qquad (29)$$

The export value here refers to total exports in the reference year and population is a mid-year figure estimated for the same year. The value of exports per capita has an obvious advantage for cross-country comparisons because the absolute value of exports may depend on the overall size of the economy and not necessarily on its performance.

14.3.2   Sector shares in manufactured exports

Just as the indicators of sector shift on the production side described in Section 3.2 should be monitored, it is also important to monitor the structural change on the export side. This shows the country's relative competitive strength in one sector over others. The computation method and classifications of exports are the same as for production but the major difference is that production data are reported by ISIC, while trade data are reported according to the Standard International Trade Classification (SITC). Whilst there are obvious differences in applying these codes, export data can be classified by the manufacturing sectors from which the products originate.

14.3.3   Export concentration index

This is the application of the well-known Herfindahl index in measuring the export concentration/diversification of a country.

$$H_j = \sum_{i=1}^{N} s_{ij}^2 \qquad (30)$$

where:

H_j = Herfindahl index for j-th product

s_{ij} = share of i-th country in total export of j-th product

N = number of countries to which j-th product is exported

The index ranges from 1/N to one (or to 10000, if the percentage shares are used). A higher value of the index indicates that the exports of a country are concentrated among a few products or a few countries depending on the variable used in analysis. A higher concentration indicates greater dependency on a few export products in the external market.

For better interpretability in data analysis, the Herfindahl index can be normalised as follows:

$$\hat{H}_j = \frac{H_j - \dfrac{1}{N}}{1 - \dfrac{1}{N}}$$

(31)

where:

$\hat{H}_j$ = normalised Herfindahl index for j-th product.

For illustration, the Herfindahl index was compiled for total manufactured exports for two countries, China and India, over the 15-year period from 1992 to 2006. In this case, notations are applied to denote the following:

H_k = Herfindahl index for exports from k-th country (origin)

s_{ij} = share of i-th country (destination) in total exports of k-th country

N = number of countries to which manufactured products of k-th country are exported

In order to obtain the value of H_k, the share of the top 40 countries in the total manufactured exports of China and India was calculated and the rest of the countries were combined into a single group which accounted for around 9 per cent of China's exports and 12 per cent of India's exports.

Figure 7: Value of Normalized Herfindahl index for manufactured exports of China and India
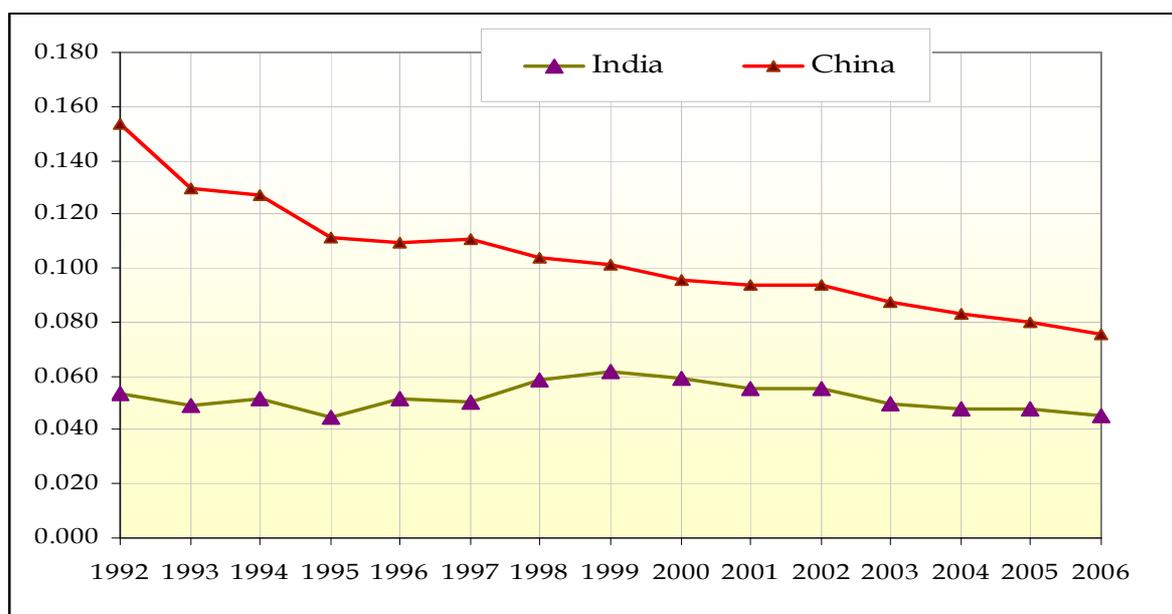
Source: UNIDO Database - IDSB

The results in Figure 6 show that for many years China's exports were limited to a few countries. At the beginning of the 1990s, the 20 top export partners had more than a 50 per cent share of China's total exports. However, China's exports rapidly diversified throughout the 1990s and significantly closed the gap with India.

15.0   Consolidated measures of industrial performance

The indicators presented above referred to the statistical measure of different dimensions of industrial performance separately. For an overall assessment of performance reflecting all

dimensions, a consolidated measure can be compiled in the form of a composite index constructed using a number of the major indicators.

In international statistical practice, composite indices are commonly used to rank countries for their achievements in social and economic development. One of the most widely used

indices is the Human Development Index (HDI)[1], which combines three dimensions of human development: a long and healthy life; knowledge; and a decent standard of living. Similarly, the Technology Achievement Index (TAI)[1] is used by the United Nations Development Programme (UNDP) to measure how well a country is creating and diffusing technology and building a human skill base, reflecting its capacity to participate in the technological innovations of the network age. The TAI focuses on four dimensions of technological capacity: creation of technology, diffusion of recent innovations, diffusion of old innovations, human skills.

UNIDO maintains a scoreboard of industrial development, which comprises several indicators related to industrial performance. A set of six indicators is taken for calculation of a composite index, which is called the Competitive Industrial Performance (CIP) index. The indicators and the underlying economic dimensions are as follows:

Table 2: Components of the Competitive Industrial Performance (CIP) index

	Statistical indicators	Economic dimension
	A. Production capacity	
1.	MVA per capita	Level of industrialisation relative to size of the economy
	B. Export capacity	
2.	Manufactured exports per capita	Export capacity and competitiveness of goods produced for external market
	C. Intensity of industrialisation	
3.	Share of MVA in GDP	Contribution of manufacturing to economy as a whole
4.	Share of MHT in MVA	Technological development in manufacturing
	D. Export quality	
5.	Share of manufactured exports in total exports	Role of manufacturing in export
6.	Share of MHT in total export	Ability to produce technologically advanced products for external market

The individual indicators listed above were described earlier. For the compilation of the CIP index, these indicators are further transformed. First, the selected six indicators of industrial performance are converted to four CIP component variables:

1) MVA per capita (no change)

2) Manufactured exports per capita

3) Industrialisation intensity *equals*

$$\frac{1}{2} \times (\text{Share of MVA in GDP} + \text{Share of MHT in MVA})$$

4) Export quality *equals*

$$\frac{1}{2} \times \text{(Share of manufactured exports in total exports + Share of MHT in total exports)}$$

The next step is to normalise the value of each CIP component variable in order to obtain the individual indices according to the formula:

$$I_{i,j}^{k} = \frac{X_{i,j}^{k} - \min(X_{i,j}^{k})}{\max(X_{i,j}^{k}) - \min(X_{i,j}^{k})} \quad I_{i,j}^{k} = \overline{0,1} \tag{32}$$

where:

$X_{i,j}^{k}$ = value of *k*-th variable for *i*-th country in *j*-th period

min and *max* = smallest and largest value in the sample

The top country in the sample obtains a 1, while the worst performing country obtains a 0. Once the individual indices are constructed for all four composite variables, the CIP index is then calculated as the arithmetic mean of their standardised values:

$$CIP_{i,j} = \frac{1}{4} \sum_{k=1}^{4} I_{i,j}^{k} \tag{33}$$

A meaningful interpretation of the CIP index can be drawn especially in the comparison of its value over time and across countries. A positive change in the value of CIP for a country over time indicates increased production capacity and improved quality of products with a better chance of penetrating international markets. The comparison of CIP across countries indicates a comparative advantage of one economy over others.

The CIP index combines measures for both the production and trade dimensions of competitiveness. However, it places major emphasis on export-oriented production rather than production for domestic demand. The performance of smaller economies with an export orientation is well reflected in the index and thus gives them a relative advantage in

the rankings. Figure 8 illustrates the ranking of countries with the highest CIP index value in 2003 compared with 1993.

Figure 8: The top 15 countries in CIP ranking

CIP index value
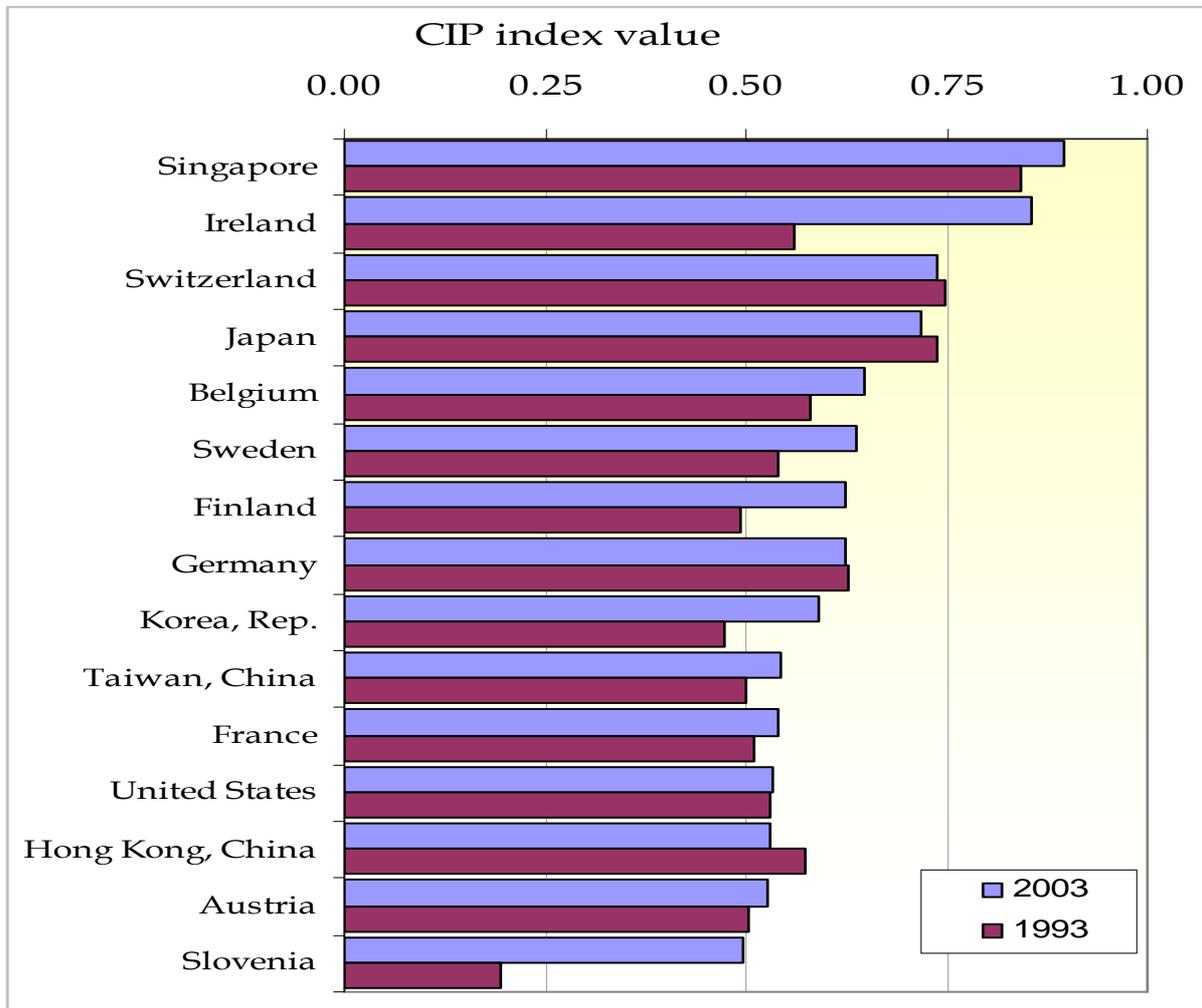
Source: UNIDO Database – Scoreboard

More export-oriented small countries like Singapore, Ireland and Switzerland appear well ahead in the ranking shown above when compared with Germany, France and the US and this may well encourage the large number of medium- and small-sized economies of the developing world to learn from the achievements of these countries in today's globalised world.

Annex 6: Industry classifications for data analysis derived from ISIC rev 3

1.   Agro-industrial sectors

Division 15   Manufacture of food products and beverages

Division 16   Manufacture of tobacco products

Division 17   Manufacture of textiles

Division 18   Manufacture of wearing apparel; dressing and dyeing of fur

Division 19   Tanning and dressing of leather; manufacture of luggage, handbags, saddlery, harness and footwear

Division 20   Manufacture of wood and of wood products

Division 21   Manufacture of paper and paper products

Group 251   Manufacture of rubber products

2.   Resource-based sectors

Division 15   Manufacture of food products and beverages

Division 16   Manufacture of tobacco products

Division 17   Manufacture of textiles

Division 19   Tanning and dressing of leather; manufacture of luggage, handbags, saddlery, harness and footwear

Division 20   Manufacture of wood and of wood products

Division 21   Manufacture of paper and paper products

Division 23   Manufacture of coke, refined petroleum products

Group 251   Manufacture of rubber products

Division 26   Manufacture of other non-metallic mineral products

Division 27   Manufacture of basic metals

3.   Energy production sectors

Division 10   Mining of coal and lignite; extraction of peat

Division 11   Extraction of crude petroleum and natural gas; service activities incidental to oil and gas extraction excluding surveying

Division 12   Mining of uranium and thorium ores

Division 23   Manufacture of coke, refined petroleum products and nuclear fuel

Division 40   Electricity, gas, steam and hot water supply

4.   ICT goods production sectors

Division 30   Manufacture of office, accounting and computing machinery

Group 313   Manufacture of insulated wire and cable

Division 32   Manufacture of radio, television and communication equipment

Group 331   Manufacture of instruments and appliances for measuring, checking, testing, navigating and other purposes

Except ISIC class 3311 manufacture of medical and surgical equipment

5.   Manufacturing sectors by technology

Low technology

Division 15   Manufacture of food products and beverages

Division 16   Manufacture of tobacco products

Division 17   Manufacture of textiles

Division 18   Manufacture of wearing apparel; dressing and dyeing of fur

Division 19   Tanning and dressing of leather; manufacture of luggage, handbags, saddlery, harness and footwear

Division 20   Manufacture of wood and of wood products

Division 21   Manufacture of paper and paper products

Division 22   Publishing, printing and reproduction of recorded media

Division 36   Manufacture of furniture; manufacturing n.e.c.

Division 37   Recycling

Medium-low-technology

Division 23   Manufacture of coke, refined petroleum products and nuclear fuel

Division 25   Manufacture of rubber and plastics products

Division 26   Manufacture of other non-metallic mineral products

Division 27   Manufacture of basic metals

Division 28   Manufacture of fabricated metal products, except machinery and
               equipment

Medium-high and high technology (MHT)

Division 24   Manufacture of chemicals and chemical products

Division 29   Manufacture of machinery and equipment n.e.c.

Division 30   Manufacture of office, accounting and computing machinery

Division 31   Manufacture of electrical machinery and apparatus n.e.c.

Division 32   Manufacture of radio, television and communication equipment
               and apparatus

Division 33   Manufacture of medical, precision and optical instruments

Division 34   Manufacture of motor vehicles, trailers and semi-trailers

Division 35   Manufacture of other transport equipment

References and Recommended Further Reading

Arrow, Jairo and Gösta Nilsson. *Progress Report for the 14th International Roundtable on Business Survey Frames*, Auckland, 2000.

Bee, Marco, Roberto Benedetti and Giussepe Espa. "A Framework for Cut-Off Sampling in Business Survey Design." *Department of Economics Discussion Paper No. 09*, University of Trento, Italy, 2007.

Cochran, William G. *Sampling Techniques*, 3rd Edition, New York: John Wiley, 1977.

European Commission/Eurostat. *Methodology of Short-term Business Statistics, Interpretation and Guidelines,* Luxemburg, 2002

European Communities. *Business Register: Recommendations Manual*, 2003 Edition, Office for Official Publications of the European Communities, Luxembourg.

Kalton, Graham. "Models in the Practice of Survey Sampling." *International Statistical Review,* Vol. 51 No. 2 (August 1983):175-188.

Karmel ,T. S. and Malti Jain "Comparison of Purposive and Random Sampling Schemes for Estimating Capital Expenditure." *Journal of the American Statistical Association*, Vol. 82 No. 397 (March 1987):52-57.

Kish, Leslie. *Survey Sampling*, New York: John Wiley, 1965.

Korns, Alex. *Updating the Sri Lankan Register of Industry and the Role of IT*, paper for the third International Conference on Establishment Statistics (ICES), Montreal, 2007.

Korns, Alex. *A Workable System for Updating Indonesia's Manufacturing Directory*, paper for the first ICES, Buffalo, 1993.

Hill, Peter. *Recent Developments in Index Number Theory and Practice*. OECD Economic Studies, No. 10, Paris, 1988

ILO/IMF/OECD/UNECE/The World Bank. *Producer Price Index: Theory and Practice*, Washington D.C. , 1994

Särndel, Carl-Erik, Bengt Swensson and Jan Wretman. *Model Assisted Survey Sampling,* New York: Springer, 1992.

UNIDO. *Handbook of Industrial Statistics*, 1992

UNIDO. *International Yearbook of Industrial Statistics*, 2009

UNIDO. *Industrial Development Report*, 2009

United Nations and Others. *The System of National Accounts*, 2008

United Nations Statistical Division. *International Standard Industrial Classification of All Activities, ISIC Rev 4*, United Nations, 2008

United Nations. Strategies for Measuring Industrial Structure and Growth, *Studies in Methods No. 65*, United Nations Statistical Division, 1994.

United Nations. *International Recommendations for Index Numbers of Industrial Production 2010*

United Nations. *International Recommendations for Industrial Statistics 2008* (IRIS-2008)

United Nations*. Guidelines on Principles of a System of Price and Quantity Statistics*, Statistical Papers Series M, No. 69, New York, 1977

United Nations. *Index Numbers of Industrial Production*, Studies in Methods No. 1, New York, 1950

Vale, Steven. *The Use of Administrative Sources for Economic Statistics – An Overview*, paper delivered at Regional Workshop for CIS on the Use of Administrative Data for Economic Statistics, Moscow, 2006

Weeks, William. *Register Building and Updating in Developing Countries*, paper for the third ICES, Montreal, 2007