

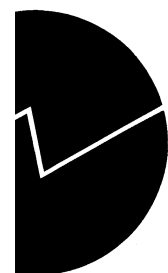
Statistics Norway
Department of Economic Statistics

Tom Langer Andersen

Documents

**Short-term Statistics in
Mozambique**

A short-term mission
19 June - 6 July 2000



Statistics Norway
Department of Economic Statistics

Tom Langer Andersen

**Short-term Statistics in
Mozambique**

A short-term mission
19 June - 6 July 2000

Contents

Executive summary.....	2
1 Introduction.....	3
2 The short-term statistical system – in general.....	5
2.1 User requirements for STS.....	5
2.2 The STS initiative.....	6
2.3 Some organisational aspects.....	9
3 The methodological areas to focus on.....	11
3.1 Population, unit and sample.....	11
3.2 Data collection – for speed and timeliness.....	16
3.3 Data editing – handling non-responses.....	22
3.4 Estimation of population totals.....	26
3.5 Analysis for quality.....	34
3.6 Meta-information – improving user accessibility to the statistics.....	35
4 A new IT-system – a strategic part of the STS initiative.....	36
4.1 Introduction – summary.....	36
4.2 Provincial Data Entry System.....	39
4.3 Input Database System.....	40
4.4 Data Administration System.....	41
4.5 Data Editing System.....	43
4.6 Data processing – population estimates – aggregations.....	46
4.7 Output Database System.....	46
5 Dissemination with a special focus on A Conjuntura Económica.....	48
5.1 The statistical calendar – advance information on publishing plans.....	48
5.2 A Conjuntura Económica.....	49
Appendix A The terms of reference.....	53
Appendix B The STS surveys of the INE – an overview.....	54
Appendix C Quality characteristics in Business statistics.....	57
Appendix D The statistical release calendar.....	59
Appendix E A documentation of the Mozambican CPI (in Portuguese).....	62
Appendix F Time schedule for the production line – an example.....	66
Appendix G Critical units in data capture and editing – an example.....	68
Appendix H The STS samples.....	71
Appendix I List of contacts.....	75
Appendix J List of Literature.....	76
Recent publications in the series Documents.....	77

Executive summary

This report presents an evaluation of the short-term statistical system (STS) of the INE. The mission has focused on the six monthly surveys, and the corresponding production lines and their efficiency. Furthermore, A Conjuntura Económica has been a special subject for studies as a part of the dissemination process.

The ambition of the short-term statistical system is to meet important user requirements. Although the current situation indicates that these will be hard to fulfil in the short run, the new surveys provide a basis for succeeding in the medium run. An important strategic target for the STS is to enable building a Quarterly National Account system in the medium or long run.

- ✓ Meetings with the users indicate that the INE must - for the next years - focus on the accuracy of the estimates, the comparability over time - and the timeliness and punctuality in dissemination.
- ✓ The STS surveys seem to cover most relevant variables needed. However, to fulfil the requirement for value, volume and price data for all sectors concerned, a revision of some of the surveys seems to be needed (especially the construction and the service sectors) where only value data are collected.
- ✓ For the STS system to stabilise - time and patience is needed. Due to several revisions during the last two years – although needed – it is not recommended to make further changes in the short and medium run.
- ✓ An immediate initiative is strongly recommended for completing the data collection part of the STS for the three first quarters of 2000. Very large unit non-response rates threaten severely the basis for producing statistics for this year - and also the quality of time series, which is important for all types of use.

An immediate action for improving the province delegation performance seems to be needed.

- ✓ The production lines do not in general seem to be set up for meeting the user requirements for speed in production and timeliness and punctuality in dissemination. Some initiatives are recommended:
 - ⇒ To develop a new IT system supporting the data collection in the provinces, the transmission (using the Internet) of data to Maputo Central, and improving data administration and quality of data processing including efficient data control facilities. A new system should be in place this year.
 - ⇒ Improve efficiency in each process based on monthly production plans (annual plans) providing clarity in tasks, priorities and responsibilities of the staff involved.
- ✓ To improve user accessibility to statistics a brief documentation model (3 - 4 pages) giving the user a quick insight into the basics of the statistics has been introduced. This type of documentation should be well adopted for dissemination using the regular publications and the Internet. A draft for the CPI has been produced and more initiatives will be taken shortly.
- ✓ A Conjuntura Económica is the main channel for disseminating the STS statistics and has been in operation for some two years. It is recommended that the publication should have a monthly frequency from 2001 on. All separate publishing of the STS statistics should be discontinued (including the CPI) and be released in A Conjuntura Económica.

1 Introduction

The visit to Instituto Nacional de Estadística (INE) – within the framework of the Twinning Project – took place in the period 19. June - 6. July 2000. The mission has focused on the short-term statistics – methodology and dissemination. See appendix A for the terms of reference.

The background for the visit is that the INE recently has launched revised monthly surveys for industry, construction, transport, commerce, hotels and restaurants and services. The surveys are still in the starting phase, and several adjustments of the methodology applied are likely to be necessary during the first year. The revision comprises for all statistics a full transition to a monthly frequency, new sampling plans and samples, and revisions of questionnaires and guidelines. A part of the background is that the survey system has been changed twice during the recent years. Some characteristics previously collected in separate surveys have been built into the business sector surveys. Some new surveys have also been initiated covering industrial sectors previously not covered.

Although the attempts made to ease the respondent burden through simplifying the surveys – the surveys still struggle with low response rates (25 - 70 per cent) – in some provinces substantially lower than this. For most surveys the lack of motivation among respondents for participating seems to represent a severe problem to the INE.

Short-term statistics are published on a quarterly basis in "A Conjuntura Económica" (ACE) which was established some two years ago as the main paper-based channel for releasing short-term statistics. Due to the response problems mentioned the publishing of new and/or revised statistics has so far only been possible to a limited extent. Preliminary results for the manufacturing sector (item statistics) based on a small sample of large establishments have been made available to the users. The INE has also initiated publishing using the Internet – so far on a small scale.

Three main subjects have been in the focus for the visit to the INE.

- ✓ The production routines of the STS have been a major topic and most all statistics have been thoroughly examined.
- ✓ The evaluation of the current systems has emphasised the need for an efficient IT system. The lack of a well-functioning IT system seems to be a severe obstacle for improvements of efficiency.
- ✓ The publication A Conjuntura Económica has been evaluated and several changes are recommended.

This report focuses on some of the important processes in the production lines of the short-term statistics and all related activities – ideas are discussed and recommendations made. The problems that the STS system recently has run into are not focused on – but serve as a background for understanding and making suggestions in a relevant way. A special focus has been on the parts of the production process where errors are likely to affect the accuracy in estimation – a key quality aspect. The quality perspective has deliberately been put in front, where some key quality elements have been focused on – as target areas to be pursued in the short and medium run. This concerns especially the need for improving the accuracy in estimation, comparability in space and over time, and the need for improving punctuality and timeliness in dissemination.

Chapter 2 provides a background overview of the current situation for the STS system in the INE. Chapter 3 looks more closely at some of the high priority issues – data collection, the processing of data (editing and estimation) and the importance of producing and making documentation available to the user. In chapter 4 the focus is on the new IT system supporting the need for an efficient administration of the data flows from the provinces. Alternative methods for improving efficiency in data editing and processing of statistics are also discussed.

In chapter 5 the focus is on the dissemination process where the need for a statistical calendar is emphasised – as a tool for improving timeliness and punctuality in dissemination. Further – a special focus is on A Conjuntura Económica – the main paper-based channel for disseminating statistics to the users. The publishing programme for the forthcoming years requires a monthly frequency for this publication – most likely to start from the beginning of 2001. All short-term statistics should be published in the monthly A Conjuntura. The INE is also recommended to continue and also speed up the process towards Internet publishing – including the publishing of A Conjuntura.

A list of persons who have contributed to this report is given in appendix E. I like to thank all for interesting presentations and discussions – and for showing me their generosity and patience. I like especially to thank the statistical advisor Leif Korbøl and his wife Inger for their support and help in manoeuvring in a Portuguese speaking environment. I also like to thank Dr. Firmino Guiliche for the social activities.

2 The short-term statistical system – in general

The STS surveys have been modernised during the last years – most recently in the end of 1999. This includes that some new surveys have been implemented as a first step towards fulfilling the ambition of building a complete system covering all business sectors. Giving high priority to the development programme has been a necessary step to meet the user requirements. This is also an important strategic initiative towards providing a basis for the Quarterly National Accounts system (QNA) – another priority area. Several meetings with users clearly indicate that the initiative is very well received.

The user meetings have also confirmed the need for an offensive initiative. The user recommendation is – briefly – to improve accuracy in estimation and the comparability of short-term statistics over time and space, and to increase the focus on timeliness and punctuality in dissemination.

Establishing a well-functioning STS requires nevertheless time – and patience. Although much has been done when developing the system during this year – much remains to be done. The main focus has to be put on the data collection process in the months to come. The respondents' motivation for providing data on a regular basis is low – likely due to a lack of understanding of the importance in having short-term statistics to monitor an economy in transition. Further – much remains to be done in the estimation process.

2.1 User requirements for STS

The user needs have in several ways – directly or indirectly – been in the focus of the INE during the recent changes in the STS system.

During the mission there have been several meetings with users. Most have been both constructive and interesting providing relevant feedback to the INE. The overall user evaluation of the recent effort to develop the STS system is positive. The initiative and the plans seem, however, to be poorly communicated to the users.

In general the user needs are numerous and vary. The focus tends – naturally – to be on the problems that the user experiences when using the INE products. The user's role in this system is – in fact – one of a critic. And let's hope that the user's role remain to be so – in a constructive way. The worst case is in fact when users and statisticians do not communicate. The INE ought to make a rapid effort to encourage the production line senior officers and leaders (e.g. annual meeting with different users) for having regular contact with all types of users.

In spite of some differences in the needs and preferences – it is possible to sum up some common key elements:

- ✓ The quality of the statistics must improve – user needs have questioned the accuracy of the estimates, the comparability over time and space, and punctuality and timeliness. For the time being these criteria are the key quality issues.
- ✓ The statistics should as far as possible – for each sector – provide a consistent set of indicators – containing value data, volume data and price data for the sectors surveyed – using absolute figures e.g. values or indices. The item statistics as released today – especially for the manufacturing sectors – seem to be less relevant.
- ✓ Most analyses require time series. Due to major revisions of the statistics during the recent years the comparability over time might have been distorted. High priority should be given and resources be allocated to linking the new STS statistics (time series) to predecessors (old time series) when available.

- ✓ Statistics are primarily wanted on a national level. Users foresee, however, a future need for STS information (as well as structural statistics) on a provincial level. This is not a high priority but will depend on the speed of the economic change of other areas in Mozambique.

The status of the STS in general and the ability to meet the needs are discussed below – within a quality framework established and recommended for quality in business statistics. For a closer presentation of the key elements in the international recommendations on quality in business statistics – see appendix C.

2.1.1 Accuracy in estimates

The evaluation of the current situation during this mission indicates that the INE needs to improve the accuracy in the estimates. The sources of errors affecting the estimates are many and occur during all processes in the production line. Some important sources are:

- ⇒ Establishing a sampling plan and the wanted coverage (sample share of population), and the selection of consistent unbiased estimators for producing population figures.
- ⇒ An appealing design of the questionnaires containing a limited number of variables to be covered, clarity in guidelines and the relevance of concepts defining the data requested from the respondent – reduce the impact of measurement errors and of course the respondent burden.
- ⇒ An efficient data collection system is of utmost importance for being able to produce good quality data at low response rates and reaching for speed in timeliness.
- ⇒ The use of relevant micro and macro controls reduces the impact from processing errors.

2.1.2 Comparability

Statistics for a given variable have the greatest value to the user when enabling reliable comparisons of values taken by the variable across space and over time. Problems with *comparability over time* might occur due to a number of causes. Changes in the statistics due to modification of the reference concept or changes in the measurement process are likely causes. The necessary links in the time series should be established (if possible) and the process should be documented, and the impact on series from the changes assessed. As concerns *comparisons across space* the typical user tasks might be to compare e.g. sales between different industrial areas, between provinces, or making comparisons with neighbouring countries.

In general, there is a need for strengthening the awareness among staff members as concern comparability. The user value of short-term statistics is closely related to the comparability over time (and space).

2.2 The STS initiative

The system contains for the time being six surveys – all revised by the end of last year and started in January this year. Some of the surveys are new while some have predecessors based on smaller samples. The new surveys seem to use somewhat different concepts for key variables including modernised questionnaires. The surveys are:

- ✓ Manufacturing production
- ✓ Construction
- ✓ Domestic trade and services
- ✓ Transports
- ✓ Airports and harbours
- ✓ Hotels and restaurants (tourism)

In addition the system comprises the external trade statistic based on administrative information and the CPI – both available on a monthly basis.

Further in this chapter some main features of the STS are presented. For more information on status and plans for each statistics – see appendix B.

The goals of the STS initiative are of course several. As a background this chapter gives a brief presentation of the main features. Some of these will be discussed and developed further in chapter 3, 4 and 5. Some important goals are:

- ✓ A full coverage of all relevant industrial sectors.
- ✓ Variables selected to cover the current activity in the business sector based on observing sales values and/or production values, volume indicators and prices. No explicit plans have been made about producing leading indicators e.g. statistics on new orders.
- ✓ Monthly data collection.
- ✓ Sampling plans are to be modernised.
- ✓ To establish efficient data collection routines – using the provincial delegations.
- ✓ All standards and classifications in line with current international recommendations.
- ✓ Establish modern IT systems supporting the methodological targets.
- ✓ Improve accuracy and comparability of the statistics – work on data editing and processing.
- ✓ Meta-information about the statistics – documentation.
- ✓ Monthly publishing of statistics – a long-run target – improving timeliness and punctuality.
- ✓ All statistics to be published using A Conjuntura Económica.

A general impression is that the STS plans provide a good basis for fulfilling these ambitions. For achieving an efficient production system a new IT system seems to be the most urgent task. An immediate action to remedy the problems that have emerged in the data collection process is nevertheless strongly recommended.

The full sector coverage seems to be established by the launching of the new STS statistics. The sectors' plans are relevant to the economy in Mozambique.

The concepts and definitions of variables adopted are developed in co-operation with the national account specifications. The variables selected are basically concurrent or coinciding indicators – for the phases of the business cycle – a reasonable choice for an initial phase. The current set of variables provides a basis for further analysis among analysts when monitoring the economic situation of the country.

The ambition of covering values, volumes and prices seems not to be fully met for all surveys during the initial phase. For all sectors and almost all variables the surveys cover value data – sales value and/or production values. Two large sectors seem to be left out in this phase: Construction and Domestic trade and services. None of these seem to cover volume or price indicators. However – parts of the domestic trade (retail sales) have the option of using the international recommended deflator produced from the CPI data. This will make it possible to produce a volume indicator for retail sales. For the construction sector the use of administrative data from the government sector might provide a basis for producing volume indicators – and even to produce leading indicators. The quality of this source has so far not been evaluated.

Somewhat more problematic in the STS system is the lack of a clear link between the plans for what to estimate (must be correlated with the expected use of the statistics) and what to observe. A seemingly minor problem – but nevertheless a complex one – is that not one of the questionnaires asks for the active period of production or sales for a month. An example may illustrate the nature of this problem.

A monthly production index will normally be used for measuring changes in the production activity and for estimating the productivity of labour – in and across sectors. For all statistics it is also important to make the months of the year comparable independent of the actual length of the calendar month (adjustments for the length of month). Some of the calendar months have 30 days; some 31 days and February might have 28 or 29 days. The average calendar month will have slightly above 30 days. The active production period within the calendar month will, however, normally be shorter than this and depend on the number of days that one works in a sector during the week. In some sectors the persons employed work 5 days a week, while in other sectors 6 or even 7 days. Some of the month to month

variation in the active production will thus be caused by variations in the number of Saturdays and Sundays from one calendar month to the next. Again – this might be corrected by introducing additional correction for the number of working days (working day adjustments). For this purpose, however, one must know the normal number of working days within each firm for a typical week. As an approximation for this one often uses general sector information when available. The types of corrections discussed are in fact required for being able to compare production (and sales as well) on a monthly basis throughout the calendar year.

The challenges of measuring production are, however, even more complicated because most units or firms might have periodic stops in production within the month, and the average number of days with stops in production for a sector will normally vary from one calendar month to the next. Making adjustments for stops in production might be of importance when one wants to utilise the production index for measuring the productivity in a sector and across sectors as well. Stops might be caused by holidays (other than Saturdays and Sundays), technical problems in the firms, strikes in a firm or a sector etc. For holidays that are common and the same throughout the country, no special treatments should be required. This is simply because the number of working hours (to be used in the estimation of the productivity) should reflect this type of stops. Other causes for stops in production where the persons employed remain in the job (doing other types of work), might require that some kinds of adjustments be made when measuring the production activity. For this purpose one has to observe the active period of production within the month. Asking the respondent to specify the periods when production has been stopped normally does this.

All surveys are based on monthly data collection routines and the ambition has been – for the long run – to publish monthly STS statistics, while the short-run ambition has been to utilise the monthly data for producing preliminary indicators on a quarterly basis. To meet the short-run ambition some indicators have been produced and released in A Conjuntura Económica. The number of preliminary indicators is expected to increase in the near future.

The adoption of a monthly survey system has nevertheless been challenged. In the initial phase there has been severe and seemingly structural problems in the communication between the delegations and the central unit in Maputo. Although not knowing exactly the actual response rates, there are indications that the rates – so far – have been low for most of the surveys. And – for some of the provinces the rates are considered as being extremely low.

- ☞ **It is strongly recommended that immediate action be taken to remedy the problems that have occurred. Close co-operation and a strengthening of the communication between the province delegations and the central unit is required.**

Existing questionnaires have been modified. Several variables previously collected in two forms have been gathered in one questionnaire to improve consistency in the data received and to reduce the total number of forms received by the respondent. The guidelines have furthermore been simplified containing the clarity needed.

It could have been argued – in advance – that initiating monthly data collection for all surveys was too ambitious. The step by step approach chosen including the intermediate targets for producing preliminary indicators on a quarterly basis makes the overall ambition reasonable.

The current use of standards and classifications does not meet the ambition to be in line with the current international recommendations. This is an important issue when working for comparability across countries. The INE has for some time been involved in international co-operation comprising the southern African countries (SADC). One of the issues brought up in this context has been to improve the comparability across countries in the region. For the classification of business statistics the INE has for some time been using ISIC, Rev. 2, which recently has been revised.

- ☞ **ISIC revision 3 is now available. It is recommended that the medium-term plans include an implementation of the new version.**

For the collection of item information in some of the surveys the INE has adopted the CPC classification. For further discussion of the classifications – see Solheim and Ragnarsson (1999).

Substantial improvements have been achieved by introducing the new sampling plan (January 2000) covering a number of sectors (2- or sometimes 3-digit ISIC). The new sampling plan is based on stratification, proportional allocation to the strata and PPS-sampling. These topics will be discussed somewhat further in chapter 3.

The data editing process including data capture must have the number one priority for the short and medium-run. The provincial task of collecting good quality data having low and stable non-response rates has to be focused on in the short-run. To improve accuracy in estimates and improve comparability special attention must be paid to the data-editing phase. Further – to improve the user access to the basics of the statistical product, meta-information i.e. information about how statistics are produced is required. These topics will be discussed further in chapter 3 and 5.

The current IT system does not meet the requirements for a broad provincial based data collection system. A new IT system supporting this and the methodological ambitions is badly needed. The new system will have a strategic role in improving data administration, give an efficient data editing system and improve the basis for control and verification of the data to be published. This and related topics are discussed and developed further in chapter 4.

For the dissemination of the STS statistics A Conjuntura has been established. The dissemination process seems, however, to require some changes when preparing for the start of the broad publishing of STS statistics. Furthermore – meetings with users indicate that A Conjuntura so far has not been able to fill the central role in dissemination as planned. Some attempts are recommended for making changes regarding the frequency and the coverage of the publication. Chapter 5 discusses this and related topics.

2.3 Some organisational aspects

From a management perspective the organisation of the production lines seems to be insufficient when reaching for an efficient short-term statistical system. Looking at the current situation in the production lines might leave an impression that the lack of a well-functioning communication system between the provincial delegations and the Maputo Central is the main bottleneck. The problems are however more complex and in fact to be found in most parts of the production lines. The organisation does not seem to operate as required when working for achieving speed in production. The targets and responsibilities might be poorly communicated – or not understood by the staffs involved. The number of tasks is many – may be too many for creating an efficient short-term production line for statistics.

An efficient production line requires a time schedule (or plan) organising the different activities to be performed during the production process. The time schedule should clarify the targets for each activity in the monthly production line and the priorities and responsibility for each of the persons involved.

☞ A production plan should be established covering all processes for each survey.

The ultimate target in time for a monthly production line is – and must be – the date when the statistics are to be released.

All activities should have a starting date (and where needed – a finishing date). The time schedule for each activity must describe a realistic progress as concerns the expected work to be done – taking holidays, vacations etc into consideration.

The plan should – at least – cover all main activities i.e. including all main activities in data collection, data editing, data processing and dissemination.

A more detailed plan might be required specifying the tasks of each staff member. Such plans should be made separately and not built into the survey production plan – as described here.

For the monitoring of the entire production process a supervisor should be appointed – having the overall response for the speed in the production line and the quality of the statistics. Each activity should have a main responsible person. The work to be done might, however, be done by a group of staff members or a single person.

Each province delegation should have a person (name, telephone, e-mail address) responsible for the data collection process for each survey. The head of the delegation should function as a supervisor and be the main responsible one.

The plan should be produced annually and be specified on a monthly basis starting in January.

A draft plan covering some months of the year – as a possible model – is presented in appendix G.

☞ **A strengthening of the competence in staff is recommended.**

For a long-term building of competence and capacity, the number of staff in the departments concerned is recommended increased by 2 persons, where competence in economics and statistics at the university level is likely to cover the most short- and medium-term urgent need. In general the strengthening of the staff is needed for building analytical capacity on methodologies, for a close and more frequent follow-up / communication with the provincial data capture units – in training and advisory activities and lastly for building IT-competence in running a new system required. The escalation of the staff is recommended to start this year. Changing the profile of A Conjuntura according to the recommendations in chapter 5 will for 2001 require competence in economics – for the editorial tasks and for developing the analytical parts.

3 The methodological areas to focus on

This chapter looks more thoroughly into six of the issues briefly presented in chapter 2. Focus is on the population, unit and sample, the data collection process, the data editing process, data processing including estimation and control of quality. The chapter ends with a short discussion of the need for a continuation of the documentation work that has been implemented.

One of the objects with this chapter has been to restrict the general discussions to a minimum. However, to provide a context or background for communicating the problems some general discussions will still have to take place. The large and principle discussions are however left out and the focus has been on giving practical advice on the many activities and tasks to be done during the main processes. It should be added that each of the main processes will have a counterpart in chapter 4 – discussed in an IT context.

3.1 Population, unit and sample

A basic part of the STS system is the task of defining the population, selecting the unit for observation and drawing a suitable sample of units from the population. The choices made during this phase and the quality of the population information are of utmost importance for the overall quality of the statistics produced. In this context the discussion takes the population information for given and concentrate on some of the other problems one might run into during the survey planning-phase.

The user of business statistics will often take for granted that the target population of a survey comprises all units of the sector or all divisions concerned with the survey. In this context the term target population comprises the units or groups of units within a sector that the survey data are intended to say something about. The target population might comprise the whole population of units in the country, all units registered in the sampling frame or all units in the latter having more than 10 employees. The main point to be made here is that the target population has to be decided by the survey planner based on the overall purpose of the statistics. Even the sampling population has to be decided by the survey planner i.e. what sectors that are to be included in the surveys for data collection purposes. In short-term statistics one often excludes the units having 10 or less employees – primarily to reduce the respondent burden. It must be added that for some sector surveys – e.g. domestic trade – such exclusions might create severe problems.

The STS initiative covers all parts of the industrial target population. The sectors included in the sampling population do not, however, include all industrial sectors in Mozambique although the overall purpose is to produce statistics about the industrial population as a whole. Although this might seem conflicting it is not necessarily so. The decision of delimiting the sampling population is based on an evaluation of the importance of non-covered sectors, which might be small in an overall population context. Whether or not to include non-important sectors will depend on their influence on the results at the intended levels for publishing, the targets established for precision in the estimates at these levels etc.

For each survey one will have a unit for analysis and a unit for observation. The enterprise and the establishment are two common units used in business sector surveys. As discussed in 3.1.2 the use of the establishment as a first choice unit for all purposes should not be considered as trivial.

The sample of units for observation is drawn using methods found in most countries – stratification for controlling the stratum variance, optimal or proportional allocation of a prefixed total number of units to be sampled, and adopting PPS-sampling. See 3.1.3 for more about this.

3.1.1 The Business Register – a joint responsibility

The Business Register plays a vital role in producing all types of business statistics – either being a part of the STS activity or when producing structural statistics. The register is the common basis for all statistics – the population of units. Keeping the register updated at all times is very important and basic for the main target of all the STS statistics, observing the changes – over time and space – in the economic performance of the population.

- ☞ **All staffs must pay special attention to the population of units. A maximum updated population based on information received implying changes in the identification of sampled units must be reported to the Business Register.**

The Business Register contains several types of statistical units – units that have been established in accordance with the national legal basis for operating in business. The enterprise is normally the legal unit and might comprise one or several establishments. The normal situation in most countries is that the enterprise contains one establishment (or the enterprise and the establishment is one and the same). For statistical purposes the enterprise should be the observation unit when information about the owner dispositions is required. For many business statistical purposes the focus is on the current production activities – a type of information expected to be found in the local activities or the establishments. For some types of variables the establishment is considered to be the best suitable unit for observation.

One should however be aware of the statistical way of organising an enterprise – into establishments – does not in general coincide with the enterprises' way of organising its departments for production etc.

- ☞ **Units defined in the Business Register should – as much as possible – have a basis in the real world.**

And sometimes this lack of consistency might be a source of misleading measurements – or measurement errors.

3.1.2 When selecting the statistical unit

All the STS statistics have the establishment as the basic unit – both for observation and for analysis. This should normally be an uncomplicated choice – especially when the industrial structure of enterprises and establishments remains more or less fixed over some time. One should however bear in mind that mergers and de-mergers which frequently occur in most countries combined with a complex industrial enterprise structure seriously challenge the traditional statistical way of thinking when selecting the unit for observation.

The analytical unit and the observation unit are two important concepts to bear in mind, when planning or making changes in a survey. The analytical unit is the unit or activity we want to study – or say something about. Sometimes it is hard or might even be impossible to observe the activity of the analytical unit directly. From practical reasons we have to choose another unit for observation – hopefully being able to provide relevant information about the analytical unit. An example might clarify the distinction somewhat:

We want to measure the production and sales when the end-produce is delivered outside a sector (where the sector might be defined as an ISIC 3-digit level group). What we in fact want to measure is thus some kind of net sector production – e.g. as an approximation for the value-added. What is produced in the sector, however, might be delivered and sold both to units within the sector – as intermediate produce being part of other types of end-use products (in other units) – and to units outside the sector. Some of the units might be part of the same enterprise, which also could be the unit responsible (here the enterprise) for selling the produce.

In this case each establishment does not necessarily know what is produced for deliveries or sold to other units in the same sector. Although some units will know, there are all reasons to expect that some do not have this knowledge. If the questionnaires are filled in there are all reasons to believe that what is reported is based on a gross production concept (and not the requested net production concept). In this case the unit for observation should have been the enterprise – likely being the only one able to provide the information requested.

This example might also illustrate some of the measurement problems that one might run into in the manufacturing production survey. The difference between production and sales (measured in quantities) for a month does not in general only reflect changes in stocks. It might also reflect production for internal use in the establishment (when producing several types of products for end use) and also within the enterprise (for distribution to and further processing in other specialised establishments). So in fact the filled in questionnaires might reflect gross production figures implying that some types of production are measured several times through the production process. Frequent variations – from one month to the next – in intermediate deliveries within a sector might induce changes in the production measurements that only reflect errors when choosing units for observation. Continuous changes in the industrial structure might have the same effect – both challenging the quality of the Business Register and the quality of the statistics.

☞ **The survey planner must be clear on what is to be measured and what unit is the suitable one for measuring the type of economic activity wanted.**

This example should illustrate the need for clarity in what is planned to be measured in a survey – and that the choice of unit (and products) to be observed matters as far as the quality of the statistics goes. On a current basis a special concern should be on mergers or de-mergers within industry. This might require a reconsideration of the unit for observation.

3.1.3 The sampling plan – current administration

Substantial improvements have been achieved by introducing the new sampling plan (January 2000). The plan includes in general all establishments (units) having 50 or more employees (stratum 5 – panel). For units having less than 50 employees the target sectors have been stratified (three strata – no. 2-4) and the size of the stratum samples fixed – using optimal or proportional allocation. For sampling the PPS-method (Probability Proportional to Size) has been used. The size of samples and populations, and the sampling probabilities measured in number of units and number of employees are presented in the following tables.

Using the PPS-sampling gives a high coverage for most surveys – measured in number of employees and at low costs during data collection (both for the INE and the sampled units). The use of PPS-sampling might however complicate the task of estimating the population totals.

The sampling plan might have some weaknesses that may create bias. This concerns especially sectors where the population structure comprises some large units and many small units – like e.g. in domestic trade and services. For stratum 2 (1-19 employees) a small number of units representing 2 per cent of the stratum population have been sampled (146 units out of 6663). These units will in practise carry a weight being slightly above 42 per cent of the total number of persons employed within Domestic trade. This is the largest stratum of the sectors concerned and twice the size of stratum 5 (the larger units). A somewhat different stratification should in advance have been considered for these sectors, e.g. increasing the number of strata and making each stratum more homogeneous. This type of problem might be treated using a post-stratification of the sampled units – to establish more homogenous strata in estimation. One should also for future purposes consider making changes in the sampling plan for these sectors.

**Table 3.1 Population, samples and sampling probabilities for the STS surveys.
Grouped by strata. Number of establishments.**

Name of survey	Size of:	Stratum ¹				All strata
		2	3	4	5	
Manufacturing	Sample	38	33	68	291	430
	Population	1243	473	153	291	2160
	Sampling probability	0,03	0,07	0,44	1,00	0,20
Construction	Sample	3	9	29	109	150
	Population	95	115	29	109	348
	Sampling probability	0,03	0,08	1,00	1,00	0,43
Domestic trade and services	Sample	146	217	176	171	710
	Population	6663	964	176	171	7974
	Sampling probability	0,02	0,23	1,00	1,00	0,09
Transports	Sample	42	62	30	80	214
	Population	318	163	30	80	597
	Sampling probability	0,13	0,38	1,00	1,00	0,36
Hotels and restaurants	Sample	75	182	50	37	344
	Population	1315	312	50	37	1714
	Sampling probability	0,06	0,58	1,00	1,00	0,20
All industries	Sample	304	503	353	688	1848
	Population	9634	2027	438	688	12787
	Sampling probability	0,03	0,25	0,81	1,00	0,14

¹ Stratum grouped by number of employees: Above 50 (5), 30 - 49 (4), 20 - 29 (3) and 1 - 19 (2)

**Table 3.2 Population, samples and sampling probabilities for STS surveys.
Grouped by strata. Number of employees.**

Name of survey	Size of:	Stratum ¹				All strata
		2	3	4	5	
Manufacturing	Sample	157	557	2618	62482	65814
	Population	5031	7963	5832	62482	81308
	Sampling probability	0,03	0,07	0,45	1,00	0,81
Construction	Sample	16	167	1098	21837	23118
	Population	491	2137	1098	21837	25563
	Sampling probability	0,03	0,08	1,00	1,00	0,90
Domestic trade and services	Sample	560	3495	6568	27786	38409
	Population	23868	15033	6568	27786	73255
	Sampling probability	0,02	0,23	1,00	1,00	0,52
Transports	Sample	183	1034	1146	48811	51174
	Population	1322	2717	1146	48811	53996
	Sampling probability	0,14	0,38	1,00	1,00	0,95
Hotels and restaurants	Sample	311	2919	1854	3436	8520
	Population	5021	4927	1854	3436	15238
	Sampling probability	0,06	0,59	1,00	1,00	0,56
All industries	Sample	1227	8172	13284	164352	187035
	Population	35733	32777	16498	164352	249360
	Sampling probability	0,03	0,25	0,81	1,00	0,75

¹ Stratum grouped by number of employees: Above 50 (5), 30 - 49 (4), 20 - 29 (3) and 1 - 19 (2)

See also appendix I – table 1 and table 2 – for more details about the sector level for each survey.

Another issue of importance when working with the domestic trade and services is that the impact of the smallest units in domestic trade sectors normally tends to vary much during the business cycle. This is due to frequent "deaths of establishments" during the downswing of the economy and "births of establishments" during the upswing. In general – this will often be the case for sectors where the costs of

entering the markets (starting a new establishment) are low. This problem might be complicated to handle on a current basis due to the information about deaths and births of establishments in total normally will be fragmented – and not available at the time of estimation. The INE must pay special attention to these sectors.

3.1.3.1 A permanent dropout – what to do?

In most all surveys one often experiences that a unit non-response might turn into a permanent dropout. The causes for a dropout might vary – the loss of motivation among the respondents, or simply a shutdown of the unit concerned etc. In the STS the methodological challenge is how – and which unit to include as a substitute.

A frequent used method is – when sampling in the first place – to draw a number of additional units for each sector and stratum and province. When a unit is confirmed being permanently out – normally a new one should be included – starting from the top of the list of units not yet surveyed.

This type of list should be updated frequently (at least once a year and more frequently when needed) and made available to the delegation offices for use in the case of having a dropout. A list delivered by the central unit in Maputo for this purpose is an important tool for checking the substitution of units made during data collection.

Due to the fact that such a list does not exist the INE is recommended to re-draw the samples for this purpose only including a number of additional units. The number of additional units must be large enough to cover the needs for likely replacements – at least for a year. All units already being sampled are retained while units not co-operating are recommended replaced. Several issues must be considered.

☞ Should a replacement unit be drawn from the same province as the one having a dropout?

The answer depends on the use of the provincial information collected. If the data are used for producing provincial statistics a supplement must come from the same province. If the sample is solely drawn to produce national statistics a replacement should be used to improve the national coverage.

☞ Should at all a replacement be made when a permanent dropout is caused by a shutdown reducing the national production capacity within the sector and stratum i.e. a real reduction of the sector and stratum population?

In principle a replacement should not be made. However, at the time when the replacement is considered – one will normally not have the information required to decide whether this will end up being a real reduction in the national capacity or not. A usual experience is that many shutdown units are restarted after some time.

3.1.3.2 A system for distributing the respondent burden – rolling units out of the sample.

In order to achieve a spread in the respondent burden over time it is recommended to include the sampled units (50 or less employees) for a fixed period – e.g. for 3 years. A unit having served the 3-year period will then be left out of the sampling frame for another 3-year period. The larger units (stratum 5), however, should be kept in the sample for as long as they are in operation (panel).

☞ It is recommended that the INE establish a system for distributing the respondent burden for all sectors where this is possible.

This requires a system for administration of the units as concerns the time participated. This should be a province delegation task. The rules for rolling units in and out of the sample must be established by the central unit in Maputo.

3.2 Data collection – for speed and timeliness

The short-term statistical surveys are frequent and repetitive. The data required in the questionnaires should after a brief introduction period be well known to the respondent and should be easy to provide. Typical for the monthly statistics is that the data collection phase is short. In fact it is normally measured in a few days. The efficiency of the data collection system relies heavily on the organisation and the competence of the staff involved and the reliability of the communication systems (postal, electronic) – and of course having respondents motivated for the task. Some of the problems that the INE has run into during the initial data collection phase are probably due to problems in the communication system.

3.2.1 The legal basis

The legal basis for production of statistics in Mozambique is a separate act on statistics, which also provides the INE with a tool for fining non-respondents. Most countries utilising this tool have experienced substantial increases in the response rates (95 per cent or more for most short-term statistics in Norway – previously a 50 - 70 per cent level). For a discussion of the issue – see Solheim and Ragnarsson (1999). I have included the last paragraph from the report as a summary of their discussion:

"In the long run, this action of strengthening the laws and the compliance with the laws are inevitable strategies if the INE shall be able to increase the response rate. The laws on statistics could also be used for tapping other administrative data sources for crucial data to be used in the production of statistics."

3.2.2 Motivation of respondents

Motivating for participation in a statistical survey is a complex task and definitely dependent on the general understanding in the respondent society (business) of the need for – and importance of statistics. The society is however not a homogenous group – as concerns personal competence of each respondent, type and size of the establishments etc. For some of the establishments – the large ones – there are all reasons to believe that the reporting task should be well known. The large units are or have been participating in such surveys from the INE or other government agencies – some of them for a long time. The medium size units might however be unfamiliar with the contents (the questionnaires and the concepts) and the requirements and expectations when having the role of being a respondent. Especially for the medium and smaller units the INE has to take the respondent burden seriously – with all the implications and meanings of the expression.

The term "respondent burden" is obviously closely related to the number of questionnaires that a statistical unit receives. A typical respondent experiences a number of requests for information where normally some are voluntary but most are not. The origin for a request received might vary, but some of the sources are well known to most establishments: The tax authorities, ministries, custom administration and other governmental institutions, branch organisations and the INE.

Normally the INE contribution to the total respondent burden figure should not be high when compared with other governmental institutions. The respondent burden and how it is perceived are, however, formed in many ways. Some respondents dislike the number of times a questionnaire is received (monthly questionnaires are received 12 times a year). The persons responsible for filling in the forms in a unit are often the same (or some few persons in the unit staff). The task of filling in the forms could be numerous in the larger units – tasks that do not necessarily give credits to the person responsible. The INE has a broader approach than e.g. ministries by including smaller units in the samples, where the capability and most likely the competence for filling in the form are more limited. The critics from the smaller units are often related to the INE delegation's ability to provide assistance when help is required, and to the INE motivation for the need and importance of giving more reporting tasks to the unit etc.

The list could have been made longer but the main points should be clear. To increase the respondent burden requires that the INE – like all other government institutions – has the ability to understand the problems of the respondent and to provide reliable signals indicating a will to ease the total burden. Showing the will to provide assistance and contributing to reducing or limiting the burden might be very important when creating understanding for the task of reporting or filling in the forms.

- ☞ **All medium or small sized units are to be visited by a delegation representative – before starting to fill in the questionnaire.**

Ask for the name of the person responsible and a telephone number to reach this person. Provide the respondent with – in writing – the name of the INE representative and the telephone number to the delegation office. This must be – as far as possible – a telephone number reaching the representative having made the first visit.

During the visit the representative should first of all provide assistance on the task of filling in the form – explaining the concepts and definitions used. The purpose is to assure that the filling in task is done correctly from the very first time.

Give information about the purpose of the survey and the importance. Give information on the period (in years) that the firm must participate.

The frequency of the survey and the last date in each month for returning the completed form must be made clear to the respondent. The respondent should receive a note – in writing – presenting the deadlines for the current year (an annual time schedule).

How the questionnaire is to be returned (delivered to the delegation office directly or collected by the delegation office) must also be made clear.

Ask to be updated when changes occur – as concerns the establishment or the person responsible.

The first visit should also provide information to the INE concerning the type of production activity of the establishment. This might contribute to improve the quality of the business register and to clarify whether the unit should participate in the survey or not. The visit should also clarify any special circumstances expected to have influence on the regularity and the quality of the data reported. E.g. some establishments have a full stop during vacation where in fact all staff members might be on leave. This type of information is important when considering imputations for unit item non-responses.

- ☞ **For the introduction interview with the new respondent – the INE should work out a standard questionnaire.**

A standard questionnaire should be worked out by Maputo CU in co-operation with some of the experienced members of the province delegation staffs. Such a questionnaire should provide a common basis for the introduction interview and to be used by all establishments. This type of interviewing requires on the job training of the delegation staff – a type of training having large impact on the quality of the data reported.

- ☞ **A new visit to each of the establishments is recommended – after six months or when any signs of non-response occur.**

The main purpose of the re-visit is to be visible to the person responsible – and to confirm that the INE will assist in solving the problems with filling in the form.

This might be the correct time to deliver a small gift e.g. presenting how the statistics are published etc.

- ☞ **The delegations must have a list presenting the critical units in data capture (and in data editing).**

A critical unit is a provincial establishment known to have large impact on the national sector estimates. See appendix H for how to produce a list of critical units.

The list should be used as a priority list when having unit non-response. If one or several of the units on the list is not responding, a delegation representative must on the following working day contact the units concerned.

The list is – as stated – a priority list. The purpose of the list is to clarify the priorities – from a national statistical point of view – when several non-responses occur. This enables the central unit in Maputo to start working (editing) on the important units – and if relevant – to produce preliminary estimates for quick releases – within acceptable time limits.

3.2.3 Administrative information

Another way of reducing the respondent burden is to utilise all relevant information available from other sources. Some of the STS surveys have e.g. counterparts in the ministries collecting almost the same data from the larger units of the population.

- ⇒ The Ministry for the manufacturing industries has a production survey covering the 125 largest manufacturing industry establishments. The production information collected and the definitions used are almost identical to what is found in the corresponding manufacturing survey. What in fact differs is that the ministry survey does not cover employment, and wage and salary information. A major difference is that the ministry survey has close to a 100 per cent response rate within 5 days after the month concerned.
- ⇒ The custom authorities have hired a firm to collect and register electronically the information filled in on the custom declaration. The same data-entering job is in fact done by the INE sometime later at costs that are unknown. There are some differences in the preliminary results presented by the custom authorities and the results later on produced by the INE. The differences are considered to be large.
- ⇒ Mozambique has recently (last year) levied a new value-added tax on almost all types of sales from establishments. Some of the information requested in the STS surveys are very likely to be found in the tax authority databases. It is not unlikely that there will be differences in the definitions of sales value for taxation purposes and what are used in the INE surveys. There are reasons to expect that it will take some time before the new tax system and the administration will work as intended.

What these three examples have in common is that they represent potential alternative administrative sources for statistical purposes that might enable large improvements in the quality of the STS statistics – both in the short, the medium and the long run. In other words – a strategic administrative alternative to be utilised for the statistical programmes for the future. To achieve an understanding of the quality of these sources and – if needed – to influence upon the quality and contents, an active engagement by the INE staff is required. This will, however, be an investment where the gains will be to harvest in the long run.

- ☞ **It is recommended that the INE immediately start discussions with the Ministry of Industry aiming at modifying the ministry's manufacturing survey for a full coverage of all variables covered in the STS survey.**

A short-run target must be to reduce the respondent burden for the larger units covered by the ministry survey. Using the ministry data will increase the speed in data collection for the establishments concerned, and make possible a monthly production of preliminary indicators. A medium-term target might be to achieve the full responsibility for collecting and analysing data to the ministry.

- ☞ **The co-operation with the custom authorities should be strengthened. This concerns especially the quality of data – both when filled in (by the trader) and the edited data produced by the Crown Agency.**

- ☞ **It is strongly recommended that the INE rapidly strengthen the co-operation with the ministry responsible for the VAT administration. The VAT information is of utmost importance when producing all types of statistics and for keeping the Business Register updated.**

The VAT information will normally cover substantial parts of the population of units operating in business. In general – the use of administrative data must be utilised to a maximum. It is important that the INE staff has an offensive role – in a constructive way – influencing the future development in contents of the administrative VAT system to fit into the needs of the statistical system.

3.2.4 Why focus on the non-response – is it a problem?

Despite the best efforts of the survey managers and operations staff to maximise responses, some non-responses are virtually certain to occur. Non-responses have two effects on data: First – the non-response contributes to an increase in the sampling variance of estimates as the effective sample size is reduced from that originally sought. Secondly – it contributes to bias of estimates when non-respondents differ from respondents in the characteristics measured.

The STS surveys are based on samples of units – samples that deliberately have been reduced to an acceptable low level. The sampling plans have been designed to achieve an acceptable precision in estimates for the planned aggregates. To control the variance in estimates within e.g. each sector by stratum, an optimal number of units have been allocated to each stratum in the sectors and then drawn using a PPS-sampling procedure. Although much effort has been put into optimising the samples – across sectors and across strata, the size of samples is normally reduced to an acceptable minimum taking into consideration the budgets and time constraints etc.

Having non-responses means that the survey estimates for a month will have to be made without the information from these units. Thus non-responses among the sampled units from an acceptable low level will increase the sampling variance.

In addition to increase the sampling variance a unit or even an item non-response might cause a bias in the estimates. It is very important to avoid a situation where the non-response distorts the quality of the statistics – in fact this might end up in publishing results that are missing the target. An example might illustrate this:

The impact of a non-response within a sector by stratum might be indicated using the sampling probabilities. For some of the trade sectors the sampling probabilities on strata level are low (2-3 per cent for some of the strata – see table 3.1 or 3.2)). This indicates that each of the sampled units implicitly carries a weight being on average 30-45 times larger than the average size of the unit for the sector by stratum concerned. The loss of information from one unit might have a large influence on the stratum results as well as the sector aggregates.

All STS samples are of course designed for the possibility of having non-responses. Thus having one or some unit non-responses do not necessarily violate the estimation process. The problem is however that we do not know in advance which unit non-responses that might be critical to the quality of the estimates.

To avoid such effects much effort is normally put into making imputations when a unit non-response or an item non-response occurs. Producing good imputations might, however, be very complicated when scarce or no information at all is available for the unit concerned. The complexity of this task escalates when units continue to have frequent non-responses over time. In this context to continue making imputations based on imputed figures, turns out to be totally unacceptable. A basic principle for the imputation process must be that the influence of the statistical producer on the results should be limited to a controlled minimum.

3.2.5 The provincial data collection process

Although there are several sources of information when producing STS statistics e.g. administrative information, and several ways to collect the information e.g. questionnaire, electronically etc., the main source for the coming years seems to be the traditional questionnaire. Data will be collected in this traditional way from the establishments located all over the country, thus the provincial delegation offices will continue to have a central role in the data collection process.

Some improvements seem to be needed for the data collection phase – both as concerns the delegation performance but also for the central unit in Maputo and its role in this process.

An initial part of the data collection process is:

- ☞ **Meet the delegations including the staffs and discuss the production plans and any changes made. Meetings are to be held – at least – annually.**

- ☞ **Introduction interview – obligatory – with the sampled establishments.**

A main target for this activity is to build positive relations to the respondent assuring that the questionnaires are filled in and returned to the delegation before the end of the scheduled time limit. It is recommended that each unit – periodically – should receive re-visits from the INE.

The regular activities of the data collection phase in the delegation offices should be:

- ☞ **Delivering the questionnaires to the respondent.**
Activity to be finished: 5 days before the end of the month.

A very important target for this activity is to assure that all respondents have received the questionnaire and that they are aware of the time limit for returning the filled in forms.

- ☞ **Collecting / receiving questionnaires.**
Activity to be ended: 2 - 5 days after the end of month.

This activity includes the administration of the questionnaires received i.e. keeping record of the flow of questionnaires received. An analysis of the not responding units should be made.

Respondents normally being delayed must receive a reminder by telephone – before the last date for returning the completed form.

Entering of the data received from the respondent is also a part of this activity. It is strongly recommended that data entries should be done shortly after having received the questionnaire. Known from experience, one will during this activity run into the item non-response problem, when having a blank cell where a zero or a positive figure is expected. In general all cells in a filled in questionnaire should contain a zero or a positive figure. The reason for stressing this is that a blank might indicate both a zero value as well as "the respondent does not know or have the true value at the time of filling in the form."

In general there will, however, be one or several item non-responses in some of the questionnaires received. This requires an action from the delegation staff. An item non-response among critical units should initiate new contact with the respondent checking the true contents of the blank cell. The amount of re-contacts must however be decided within a budgetary context.

The main target for this activity is to achieve high response rates and good quality of the information received.

- ☞ **When having non-responding units – a rapid re-contact is needed (by telephone or visit if considered necessary). The critical units should have the first priority in the initial phase of re-contact.**

Activity to be ended: 5 - 6 days after the end of the month.

An important target is to collect all information from the critical units.

- ☞ **The data entry for non-responding units – control of data – using the provincial application for this purpose.**

End of activity: 6 days after the end of the month

- ☞ **Further re-contact with the respondents due to non-response or anticipated errors in the questionnaires received.**

- ☞ **Data to be sent – by e-mail – to the central unit in Maputo.**

Last date for the first transmission of data: 7 days after the end of the month.

The data sent to Maputo Central should be split in two parts or files. The files should contain records based on a standard record format including e.g.:

- the survey concerned (survey code),
- the unit (id code)
- all data elements to be collected for the survey.

The survey code could be e.g. 01 for the manufacturing survey, 02 for the construction survey etc. The id code should for each unit be identical to the establishment code used in the business register.

The first file should contain information about the sampled units or establishments – with a special focus on units having changes in relevant characteristics – for the upkeep of the national business register as well as the survey register. This type of information will also be of utmost importance for the handling of non-responses later on in the production process. The second file should contain survey data comprising all data collected. When having non-response additional information – this should be a part of the file.

The standard record format to be used in data transmission from each of the provinces must be a built-in part of a data extraction application – being a part of the provincial data entry system.

For units from which no questionnaires have been received the unit code – single – should be included in the extract from the province database. The unit code indicates that there is a non-response for the unit concerned. Text or codes explaining the cause for the non-response should be a part of the record – when additional factual information about the non-response is available.

- ☞ **If a province still have non-responding units, further re-contacts should be made. Data from questionnaires received are to be data-entered and mailed to Maputo.**

Final date for transmitting data: 12 days after end of the month.

What has been drafted above is some kind of activity plan providing information on the deadlines for finalising each activity. The dates are specified to indicate the speed required to achieve an overall target for producing and publishing national statistics. The dates used are of course tentative and have to be fixed by the Maputo central to fit into the national target for producing and publishing these types of statistics.

The activity plan also provides a timetable for each activity in the monthly production line i.e. a timetable for the production line. In this example the activities related refer to the data collection phase. See appendix H and chapter 2.4 for a more general timetable covering the whole production line.

3.3 Data editing – handling non-responses

The data editing process might contain several tasks. In fact – in some contexts the term comprises all types of activities to be done from data capture until the complete production base or file is finished. In some cases one also includes the subsequent data corrections made based on the macro controls.

In this context the term data editing will be used for a sequence of activities (or processes) starting when the data from the firm are data-entered (or optically read etc.) and stored in the provincial databases. The data editing process continues until the file or base to be used in production is finished.

Although the organisational aspects for the data editing process as such are not a part of this chapter some minor reflections will be expressed (but not pursued later on). Using the delegation offices for the data capture and the data entering processes urges for a decision on which organisational unit(s) that shall have the key responsibility for the data editing processes. Although the province delegations must have a key role in this process – simply due to the closeness to the respondents and the local knowledge – one can not ignore that the survey plan is designed and data are primarily collected for producing national aggregates. This dilemma must be handled in one way or another.

3.3.1 Data administration

This part of the system focuses primarily on the data administration in Maputo Central. As a deliverer to the central system the provincial routines play an important role. Chapter 4 sketches out the applications to be used in the provinces and the one to be used in Maputo Central.

Both the provincial and the central systems comprise data administration in the sense that is discussed above. Due to this – the chapter will primarily focus on and describe the central routines.

☞ Data to be transmitted to Maputo Central

The data to be transmitted to Maputo Central will be an extract from each of the provincial databases. The structure of a typical record will depend on the type of information that is collected.

In principle a transmission should be divided into at least two separate parts or files. Each file should have a standard format for this type of transmission and should also be a built-in part of the data extraction application in the provincial data entry system.

The first file – and each of the records transmitted – should contain all relevant information about the unit – id code, name, address etc. The focus must be on the changes in population i.e. whether basic information about one or several of the sampled units is changed.

And a second part containing the survey code, the unit id code and all data that are to be collected.

For both files) the record should be structured according to a standard fixed record format decided for use in the transmission between the provinces and the Maputo Central.

All data should automatically be loaded into the correct database. The first file should be loaded into the data administration system base and the second into the input or raw database. During the loading of data to the databases concerned, some checks should be made – see below.

When data have not been received from a province delegation (+ 2 days after the scheduled date for transmitting data) – the delegation head is contacted.

☞ **Checks for doublets – automatic.**

The purpose of this check is to eliminate any doublets in the data files received from the provinces. While loading the two files (part one and two) an automatic check for doublets should be made.

If doublets are found in a transmission the set of identified errors (doublets) should automatically be loaded to e.g. an e-mail and transmitted to the province concerned – for control and correction or possibly verification.

Data passing the check for doublets are loaded to the raw database in Maputo Central.

☞ **Checks for non-responses (item non-responses and/or unit non-responses).**

The purpose of this check is to identify unit non-responses or item non-responses. For this purpose an operational definition of non-responses must be established. A possible definition could be:

- For a critical unit (respondent) – if less than 50 per cent of the variables asked for in a questionnaire are non-blanks a unit non-response is found.
- For other units – if less than 25 per cent of the variables asked for in the questionnaire are non-blank a unit non-response is found.

This check should utilise all types of supplementary information received from the provinces explaining such non-responses.

A non-response among critical units (according to the list of critical units produced for each of the provinces) and defined in the sense above – should generate an automatic report back to the province delegation urging for a new contact with the unit(s) concerned. For all critical units the set of data must either be complete or explained by supplementary information added to the record.

When having a re-contact with the critical units the province delegations must ask for some tentative figures (when no complete information is available) to establish a basis for making imputations!

All checks are to be automated.

Tables mapping the non-responses – on a province by sector level, and at a national level – should be produced for analysis. Non-responses within the group of critical units should be presented separately. For the other – non-critical units – the table provides a basis for deciding on further action – contacting any of the provinces for further data capture / clarification – and/or continuing the MC part of the production process.

All province delegations should on a monthly basis receive the same table. An additional table should show a rank of the delegations based on their ability to achieve high response rates.

☞ **Checks for large deviations.**

The purpose of this checks is to identify large deviations in the growth rates on unit or unit by item level (cell inn the questionnaire). Large deviations should – when correct – have been commented on in the transmitted file from the province to Maputo Central.

- ☞ Automatic comparisons to be made during loading for the item response i.e. separately for each variable by unit. The purpose of this check is primarily to clarify that no records should carry errors that should have been identified and corrected by the province delegations.

Such checks should be made in the time dimension e.g. comparing this month's figure for a unit by item with the value of the preceding month and with the item value in the same month last year. Both growth rates are to be measured in per cent.

If deviations exceed the in advance fixed accept-limits (a range) – the records concerned should be reported back to the province concerned for checks and corrections or possibly verification.

This check depends on the availability of comparable data from previous years. For the first year no 12-month growth rates are available. Checks on data to be performed later on in the Maputo Central system should enable an identification of such errors. To allow enough time for checks and corrections in the provinces, these types of checks should preferably be performed during the initial loading of data.

3.3.2 Data editing in Maputo Central – for national purposes

In addition to the control of data received from the provinces – during loading of data, there must also be a set of controls for national purposes. For this part of the data editing process, we take for granted that all checks and re-contacts made with the provinces or the units concerned during the initial loading of data, have been successful. This does not, however, imply that all units have responded – thus there are still non-responses. Further – not all units that have responded have returned a complete questionnaire. In addition – there are all reasons to expect that some of the filled in questionnaires contain data elements, which do not fit in – in the sense that the estimates to be produced will be biased in one way or another. Thus a data editing process must be run for the purpose of producing national level statistics. During this phase one will have to consider imputations for some of these data elements.

Imputation is the process used to solve problems with missing, invalid or inconsistent responses identified during editing. Changing some of the responses or missing values on the record being edited to ensure that a plausible, internally coherent record is created does this. Some problems are eliminated earlier through contact with the respondent – from the province delegation or directly from the Maputo Central – or through a manual study of the questionnaire. It is generally impossible to resolve all problems at these early stages due to concerns with the response burden, cost and timeliness. Imputation is then used to handle remaining editing failures, since it is desirable to produce a complete and consistent file containing imputed data.

Although imputation can improve the quality of the final data by correcting for missing, invalid and inconsistent responses, care should be taken to choose one or a set of methods that do preserve the relationships between variables. The methods selected must not distort the underlying distributions.

According to the main line for the new production system sketched out in chapter 4, all data editing work in the Maputo Central should be done on the data stored in the production database. All data as received from the provinces should be stored in the "raw" database, and adjustments in this base are only allowed (for the Maputo Central staff) when new information is received from the province delegations or directly from the units concerned. All changes in the raw database done in Maputo Central should be logged. The basic idea behind separating and storing the raw data as collected is of course manifold, but two aspects should be mentioned. Firstly – to store the data as received from the provinces is vital for the follow-up process and contacts with the provinces. Secondly – the raw data will when compared with the later on finished production data (for a month or sequences of months) enable an analysis of the Maputo Central impacts on the data editing phase. Having the opportunity to do such analysis is important when studying the quality of the work done and the use of resources spent during the separate phases of data editing.

The following points describe a sequence of activities to be run during the data editing process for national purposes in Maputo Central.

☞ **Loading to the production database.**

The raw data from the provinces should automatically be loaded to the production database – as the correct data have been received. Any changes in data received concerning previous months are also to be loaded. This might include new data for units that in previous months have had total non-responses or item non-responses.

New information for previous months might imply that previously published results should be re-estimated and revised results for the months concerned published.

☞ **Analysing the production database.**

For the analysis of the production data – a set of tables should be produced to evaluate the current contents of the production database. The overall purpose of this analysis is to decide whether the data available will produce unbiased estimates on the national level. In this context the national level should be understood as the total aggregates as well as the sector aggregates for the country.

The anticipated impact from units missing must be evaluated and a decision made whether missing values should be imputed. For this type of analysis additional information from the province delegations concerning causes for non-responses is of utmost importance. If a small or a large unit in a province has stopped the production in a month due to vacation – and this also is the reason for the non-response concerned, this type of information must be used when making decisions on imputations. In general this type of non-responses should not be imputed. Using zero production will in fact match with zero hours worked (when estimating the productivity) and also reflect the typical seasonal patterns in production due to vacations etc.

The analysis procedure is closely related to the next operation – identifying extremes.

☞ **Identifying extremes on a national level.**

Identifying extreme values (anticipated errors) in the current month's data – having the focus on national sector level. The IT system method for identifying extremes should be based on well-known methods in use for this purpose. Such checks are somewhat more discussed in chapter 4.5.1.

These checks should be done on adjusted sets of data i.e. adjusted in the sense that all types of non-responses are excluded. The main purpose of these controls is to identify extremes (values, change rates etc) i.e. values that do not fit into the underlying statistical distribution of the adjusted set of data. For STS statistics for which publishing on the sector levels will be common, such methods will often focus on the distribution of changes over time for the group of units (sector) or the distribution of e.g. value data in the sectors concerned.

For some surveys one can utilise combinations of variables e.g. production value per employed person – in time and/or the cross sector dimension – in order to identify extremes. In such checks a closer look on each of the variables concerned will be needed to identify which one is the extreme. Such combined checks might however fail when facing double errors neutralising each other.

A suspected error might occur in one or several of the dimensions being in the focus of this check, and all types of errors should automatically be coded (flagged) by the IT system. Normally – errors flagged in several or all dimensions are to be treated as being more complex than errors only flagged in one dimension. In general – the importance of a flagged error must be evaluated in the context of which the results are to be published. If the publishing focus on the changes over time – anticipated errors in the time dimension must receive special attention.

When results are to be published on sector level (and the sector changes over time) the routine for identifying extremes should work at the sector level changes – identifying the units responsible for the anticipated errors distorting the sector performance.

Several variants of extremes might occur – in a single variable value, sum of values for single variables etc. How these should be handled must be defined in advance and some might be handled automatically by the IT system (e.g. special editing routines for handling errors in the sums of values).

☞ **Handling extremes – capacity.**

The editing capacity will normally be limited. In some months only few extremes occur while in other months the number might be large. The MC staff capacity to handle extremes should be defined in advance, e.g. limited to *nn* extremes e.g. 50.

If more than this is flagged by the system (as extremes) – the list of 50 should contain the 50 errors – ranked (top-down) – expected to have the largest impact on the national aggregates. Other extremes – not represented on the top 50 list – should be handled automatically by the IT system in accordance with predefined rules for handling such errors. This could be a correction based on hot-deck principles like:

Replacing the extreme values with a specific donor value from the sector sample of non-flagged units or the mean value of the other non-flagged units for the sector concerned.

Or:

Replacing the extreme with the last month's value for the data element concerned adjusted with the changes in the donor value from – or the mean value of – the other non-flagged units for the sector concerned.

Such corrections could also be based on cold-deck principles using last year's information for the unit or unit by item concerned.

A third approach could be to leave such corrections out for the estimation process when using an estimator like the Horvitz-Thompson estimator (or other estimators) by introducing adjustments to the inverse sampling probabilities (calibration). See chapter 3.4 for some more about this.

For the micro controls the Maputo Central staff follows up the list of 50 extremes. During this phase the staff might have re-contact directly with the units concerned – or indirectly using the province delegation responsible. A reminder – all changes based on new information from the unit should be entered in the raw database and then automatically loaded to the production database.

☞ **The production database is ready.**

When the extremes are handled – the production database should be considered ready for a first preliminary run – producing the sector aggregates.

When later on analysing the outputs of the estimation process – the sector or other aggregates in a time dimension – one will find performances in the aggregates indicating what could be non-identified micro errors. For the sectors concerned a revision of sector micro data might be needed before a final production database is prepared.

3.4 Estimation of population totals

Estimation is a process that consists of assigning values to unknown population parameters by using information from a data set. The parameters that are to be estimated can mostly be expressed as functions

of the population totals. Examples include simple descriptive statistics such as totals, means and ratios, as well as more complicated analytical statistics such as regression coefficients. Measures of precision are usually associated with these estimated parameters. While the quality of the computed estimates largely is dependent on the preceding survey steps, the choice of estimation method can also play an important role, in particular when appropriate use of auxiliary data can be made.

Good estimation methods will yield design-consistent estimates when there is no non-response. The basic consistent estimator is the Horvitz-Thompson estimator. It weights data with weights equal to the inverse of the inclusion probability generated by the sampling design. Such a weight is called a sampling weight and can be thought of as the number of times that each sampled unit should be duplicated to represent the full population. At times, there is a need to adjust the sampling weight, for example, to adjust for non-response or incorporate auxiliary information to improve the precision of the estimator. The incorporation of auxiliary information in the estimation process is known as calibration, and the resulting factors are known as calibration factors. The products of the calibration factors (adjusted for non-response) and the sampling weights are known as calibration weights. These calibration weights incorporate the sampling design with the auxiliary data. Such adjustments will generally result in estimates that are approximately design consistent, and the bias is usually negligible. The proper estimated variance of the resulting estimator reflects both the sampling design and the use of the auxiliary data.

3.4.1 Data processing – what to estimate.

For most surveys the goal for the estimation problem will normally be decided in an early stage of the planning phase. The design of a survey (size of sample, how to collect data etc.) is based on a decision on what results that are to be produced i.e. the sector levels, types of aggregates etc. This is normally based on an evaluation of the most typical users (national accounts, ministries, professional customer etc.) and their needs. A part of the planning phase is a discussion on how to estimate (or assign values to) the unknown population parameters by using information from a data set.

The remainder of this chapter is based on the assumption that the overall purpose for each of the STS statistics is to produce population estimates for the whole population of units in the country with a main focus on the ISIC sector aggregates and totals.

As discussed briefly in the beginning of this chapter, the estimation problem will be highly dependent on the quality of data that are collected. Some of the problems related to the variation will be due to the sampling error. But – just as important is the non-sampling errors – measurement errors that might have several causes. Special attention must be paid to the non-response problem. This must be handled in one way or another – depending on the nature of the non-response. The basis for a decision on how to handle the non-response should be prepared according to the principles sketched out in the previous chapter.

The basis for the estimation process is of course the updated production database for the current month. In addition – auxiliary information about the population of units will normally have positive effects on the quality of the estimated results. Such information can be sales or production values from the annual surveys (based on definitions equal to those used in the STS surveys) for the sectors concerned. If such data are available for each unit or as totals for all units in a sector, this type of information can be used directly in the estimation process.

3.4.2 Production statistics – model

The further presentation of the estimation process is based on an example – using production statistics as a model. This should cover the basic steps to be run for all types of statistics.

Before discussing the estimation problem in more detail, some basic issues concerning the purpose of the production statistics should be discussed.

3.4.2.1 The purpose

Production statistics are most frequently produced and published as an indicator of the changes in the value-added for the sectors or the industries concerned. This way of producing and publishing production

statistics is based on international recommendations and on the need for having preliminary and rapid information as concerns the changes in the GDP (gross domestic product). Some countries have started to use the production statistics as a gross production indicator when entered into the national accounts. But – for most all other regular users the production statistics are expected to be a value-added indicator.

Production statistics are, however, normally based on gross production indicators e.g. production measured in quantities, sales values etc. This might seem contradictory when compared with the national accounts' GDP concept (value of gross production less value of input factors in production). Production statistics as a value-added indicator are of course based on an assumption that a fixed relation between production and inputs exists for short periods of time. This is an acceptable assumption for most user environments and should as such not create any problems. Studies from several countries have shown that such relations exist. In the long run the relation between outputs and inputs must be expected to change due to technological developments, improved skills in staff etc.

Production statistics based on this assumption adjust to the overall purpose in several ways. Firstly – in using value-added weights for aggregation across sector aggregates based on gross production indicators – and secondly to have frequent changes in weights – annually.

3.4.2.2 Two approaches – measurement errors

Production statistics might be estimated using several types of data. Two common approaches for measuring production activity are:

- ✓ Observing produced quantities of goods or other indicators for produced quantities (hours worked, quantities of inputs etc).
- ✓ Using sales values.

The first approach seems to be planned for the industry survey in the STS programme. The survey covers produced and sold quantities as well as value of production and sales. For most of the other surveys the second approach seems to be the one considered as relevant.

When measuring production activity using sales values – one will often face some problems to be resolved:

- ✓ For some sectors the sales activity and the production activity do not coincide in time. E.g. for the agriculture sector the production will be concentrated on the harvest period. The sales might however be spread out over time (over a longer period than the harvest period) depending on the storage capability of the produce concerned and the storage facilities available for the producer.
- ✓ Similar problems can also be found in industry sectors where the time for producing a product (e.g. large machines) exceeds the month. The sales value for a month will normally reflect the period for paying the product, which might differ from the period when the product was made.

For both of the problems discussed some types of adjustments for changes in stocks or inventories have to be made in the sales values. Using the production values, which are observed in some of the STS surveys might give the adjustments needed as concerns the basic products for the establishment, but might also visualise other types of errors. A zero production value might be correct as concerns the products specified in the questionnaire, but might ignore the fact that other types of production are going on. See also chapter 3.1.2 for more about the measurement problems.

- ✓ If one succeeds in making the corrections needed for changes in inventories (on a monthly basis) the use of sales values might be a well-functioning indicator.
- ✓ For the estimation of the volume of production and changes in this – one must also have a suitable price index for deflation purposes.

If one decides to measure the production activity using produced quantities, some of the same problems as well as others might be found:

- ✓ When observing products and the quantities of these, the definitions of the products might be of vital importance. Broad definitions might cover large variations in the product qualities that might vary much from one month to another, i.e. the implicit price of the average product might show large variations from one month to the next.
- ✓ Measuring production in physical units like number of items produced, litre or gallons, kilo, tonnes etc might seem trivial, but is definitely not so. Measuring the production of beer in litre will normally be a close approximation to what one intends to measure. If one wants to observe the production of machines, the use of physical quantities might turn out irrelevant. One will often experience that the respondent expresses doubts on what is measured as well.

The basic problems discussed above focus on the measurement errors that might occur. These errors will be highly correlated with the way a product is measured and how the product is described or defined. The basic problem in this is met and understood when estimating the implicit price to be used in estimation. See also chapter 3.1.2 for more about the measurement problems.

- ✓ A well-known problem in surveys based on products is that the respondent forgets or does not understand the importance of introducing new products when the old ones are losing their importance. This type of problem requires frequent analysis to identify establishments that show a decline in the estimated production value. If such problems are allowed to develop, the impact of such measurement errors might become large.
- ✓ As mentioned above the problems related to time-consuming production are found in industry sectors, i.e. the production time of a product (e.g. machines) exceeds the month. For such sectors a somewhat different approach might be needed during the data collection phase. Replacing production measured in physical quantities with other indicators like hours worked – should be considered.

Again this is a type of measurement error that must be considered by the statisticians. Large variations in the implicit prices of a product might indicate that there are measurement problems that should be handled.

3.4.2.3 Production statistics – the estimation

The estimation of production statistics might be done in several ways and as discussed above be based on at least two types of data.

A common model

This model is often found and utilises three basic types of information:

- ✓ **The current set of data based on produced quantities or other physical indicators**
This part has already been discussed in detail in this and other chapters and should not need any further discussions.
- ✓ **A set of base prices for each of the products observed**
The base prices refer to a set of prices for each of the products observed during data collection. The prices should be estimated using a set of base period prices. The last year's average prices are frequently used for this purpose.

These prices are to be used when converting the current month's observed produced quantities for each unit by item into values measured in previous year's prices.

This enables adding across all the products belonging to one sector as well as adding for each unit belonging to a sector. Such operations would otherwise have been impossible simply due to the variations of physical units. Through this operation each of the products is assigned an implicit

weight reflecting the product's contribution to last year's total production value. And further – this operation enables comparisons between this year's monthly data and last year's data.

- ✓ **Annual structure survey data for value-added measured in factor prices**
The estimate of value-added at factor prices on sector level is normally produced in the annual structure surveys and is a close estimate for the value-added concept used in the national accounts. See chapter 3.4.2.1. This set of sector data will serve as weights when aggregating from the sector level estimates to the national level or for all other cross sector aggregates.

The main estimation problem is to produce reliable sector production value data measured in last year's prices using the sets of sector data made available through the monthly surveys.

Two basic types of statistics should at least be produced.

- ✓ Statistics showing the basic production without making adjustments for stops in production during the month, no working day adjustments should be made.

These statistics should simply reflect the gross production as it has been observed. E.g. for a constant intensity in production during some months the production should be larger for months having 31 days than months having less days than this.

This is some kind of raw statistics that first of all will have professional users. It might also be an alternative for the national accounts. Series from these statistics will however be hard to use for the regular user simply because it is complicated to see to what extent the variation in raw production is due to variations in the number of production days etc.

- ✓ Working day adjusted production statistics where adjustments have been made – for periodic stops in production, for variation in the number of working days etc. For these statistics each month will have a standard length making direct comparisons – across months as well as across sectors possible for the users.

Assuming that the set of last year's prices and the weights are produced, a typical production process will contain the following sequence of activities:

- ✓ **Estimate the production value in base year prices.**
After having checked each of the records in several dimensions, this should be a straightforward operation.

What to estimate is the production value in base prices. For each product one multiplies the number of items produced by the corresponding base price. Note that the base prices may vary across the units within a sector having the same products – defined according to the CPC product classification. When observing such variations in prices this should be utilised in the estimation due to the fact that the variation might reflect quality differences.

It is recommended that the estimated production values for each unit by product are stored in the regular production table / base as new (additional) variables.

- ✓ **Adjustments for stops in production.**
An important part of the estimation process is to make adjustments for stops in production during a month. This could be stop due to technical problems, periodic lack of raw materials etc. Most such stops will normally be short without forcing the unit management to put the staff on temporary leaves. An example might clarify this:
- ✓ *Let us say that the number of items produced in January is 100 – for a unit having only one product. The number of working days is 23 (assuming that the unit staff works 5 days per week). The unit has informed the province delegation's data collector that for 2 of the days in January there was no production due to technical problems.*

This should not affect the raw production statistics, but some additional adjustments are needed for the working day's adjusted statistics. As long as this is something that concerns the individual unit and not the sector, the adjustments needed are to be done on the data received from the reporting unit. A simple adjustment factor (the unit factor) should be made to adjust the raw production value (price multiplied by the number of units produced in January) by the unit factor. The unit factor should be estimated dividing the common number of working days for the sector (23) by the actual number of days that is worked (21). This gives in this case a factor of 1.0952. What is actually done is to estimate the production for the unit concerned assuming that there in fact were 23 working days having the same intensity in production.

Information about such stops must either be collected as a part of the questionnaire or as additional information from the unit concerned. It is of utmost importance that such information is forwarded to the Maputo Central – as a part of the regular data transmission. The estimation of the special correction factors could be automated when the information needed is reported in a numeric way (and according to a standard format).

It is recommended that these adjusted estimated production values for the units and products concerned are added to the regular production table / base as new (additional) variables.

The number of units for which such adjustments should be needed will normally not be plentiful – but might be found more frequently during some parts of the year and in some sectors more frequently than in others. Such stops might in some months be large enough – measured in days – to have an impact on sector level estimates.

✓ **Working day adjustments.**

In addition to the special corrections discussed above, some further adjustments are required for making the production statistics understandable or accessible to the user. This is the working day adjustment.

The working day adjustment operation could be done directly on the production value estimates at the sector level for a month. For making the adjustment on the aggregate level, it must be a reliable assumption to make that a common (standard) number of working days per week are found for all units in a sector. If no such standards are common the working day adjustment operation must be individualised and might become complicated due to the fact that individual data for each unit will be required. The example below is based on a standard assumption and discusses a case for a sector having 5 working days per week.

For a standard year the number of days is 365 with some 52 Saturdays and some 52 Sundays. The total number of possible working days is thus 261 i.e. $365 - 52 - 52$. The average number of working days per month for the example sector is thus 21.75 i.e. $261 / 12$. In January the total number of days are 31 and let us assume that there are 4 Saturdays and 4 Sundays. The total number of working days in January is thus 23 i.e. $31 - 4 - 4$. The correction factor needed is estimated by dividing 23 by the average number of working days per month i.e. 21.75. This gives a factor like 1.0575. Estimating such factors must of course be made for the remaining months of the year before the set is complete. The average factor for year should be equal to 1 or very close to that.

We notice that the factor depends on the actual number of Saturdays and Sundays in each month as well as the total number of working days in the year. Due to the fact that the number of Saturdays and Sundays in a month might vary from one year to another, a new set of factors must be estimated for each year.

If some of the sectors having 6 or 7 working days per week, the correction factor must be calculated separately for those sectors. In the case of 7 working days per week no working day adjustments are in fact made. The resulting factors will only reflect the variation in the total

- ✓ number of days compared with the average number of days per month. Thus, it will only be an adjustment to a standard length of the month throughout the year.

For making the complete correction for the variations in the number of working days, other days where production is stopped must be included in the calculations. Some of these stops might be due to public holidays or other stops – closely related to the annual calendar. When such stops are known in advance, they may be included in the regular working day's adjustment factor estimation. The estimation procedure will be the same, but more days are to be deducted in one or several months, in the total number of working days of the year and in the subsequent estimation of the average number of working days per month.

The general working day adjustments should be made later on in the estimation process – on the sector level – as long as there is sense in saying that the number of working days is fairly standardised across the units in a sector.

The set of working day adjustment factors estimated should be produced in advance of the calendar year concerned, and stored in special tables to be used in the corrections on sector level – later on in the estimation process.

- ✓ **Estimating the sector aggregate.**

The intention is to produce population estimates for the sector. If the sampled units could be assumed to give the best possible picture as concerns the changes in the population, no further adjustments should be necessary. And further – if the results only are to be published as indices, to give the user indications of the sector changes in production, the sample estimate could be said to represent the best possible estimate for the changes in the population production.

Population figures are needed.

For various reasons, however, some further adjustments are needed. For the sample estimate to be the best representative for the population parameters, one has to adjust for the share of the population of units not represented in the sample. Further adjustments should be made for the non-responses most likely existing. An example produced by the statistical advisor Mr. Korbol showing the impacts from the non-sampled units has previously been presented to the INE staff. The example clearly illustrates the impacts of assigning larger sampling probabilities to the larger strata, which is the case for all the STS surveys. This must be adjusted in the estimation process to avoid severe bias in the population estimates using stratum estimates. In addition the way of handling non-responses might require further adjustments in the sample estimates in order to make the appropriate corrections required.

Estimate on stratum level.

Due to this – the further estimation process for the sample estimates and subsequently the population estimates must start out on either the planned stratum levels or in post-stratified strata. To simplify the further presentation we will refer to strata and leave it up to the INE staff to decide whether a post-stratification might be valuable for the overall estimation process. Notice that this is independent of whether one publishes the results using indices or other indicators.

The population.

The national business register comprises all units running on a regular basis within the industry sectors covered – and constitutes the population for which the parameters are to be estimated.

Identify the part of non-response that remains to be handled.

The first step will be to identify the part of the non-responses left out for further treatment during the estimation process. The treatment of the non-responses is very important to avoid bias in the estimation of the population parameters. The list of 50 represents the extremes that have been handled manually in the initial phases. This does not, however, cover non-responses, but are extremes identified within the data set actually reported from the sampled units for the current month.

Longitudinal weights for the STS statistics.

Previously in this chapter a set of weights based on value-added data has been assumed available. If this is so, the data required for making the periodic structural adjustments to the STS statistics are in place. However, if this type of data is missing this will have implications for the estimation process.

Even if existing – such weights might be old, often reflecting a period of 2-5 years back. If the weights still can be considered as fairly representative for the structures – and also the current changes in the structures – they should be accepted. Although the structure data (weights) do not necessarily give a very good picture of the structures they should be preferred. However, if the industry structure has changed substantially (the number of units increased, new sectors that are important and not found in the structural surveys) one should reconsider the estimation process.

For the remainder of this chapter we assume that the set of weights is expected to work in an appropriate way.

Step 1.

For each stratum the sample sum of production values in base prices should be produced. The imputed values for produced quantities should be used when corrections are made. If parts of the non-responses have been imputed already, these data should be a part of the sample sum estimate. For the units having had special corrections in the production values, the corrected data should be used – not the raw data.

The variation for each stratum should be estimated based on the same set of data. These variances should be stored in separate tables and the time series studied in order to identify whether systematic processes are working on the sample estimates. If the variation seems to develop systematically over time – away from its historical trend – this might be an indication of errors in the process. If such errors are found, the estimation process should be stopped and the problem resolved.

Step 2.

If non-responses on stratum level still remain to be treated, an option might be to make adjustments in the original stratum sampling probabilities (calibration weights).

Step 3.

Adjust the stratum estimates using calibrated original sampling probabilities. This should give the estimated population totals on stratum level.

Step 4.

Add up the stratum estimates to the sector level.

An analysis of the time series estimates of a sector population should be done including the current results. Any indications of substantial deviations from the basic trend should be followed up searching for possible causes related to the production process. Bear in mind that changes are reflections of the true changes in the population. It is of course very important to include such changes when they are significant.

The steps 1 - 4 could be used for the estimation of population figures for values, physical quantities etc. as well. If no aggregate sector data (weights) are available the estimation process sketched out could still be used – utilising the inverse sampling probabilities adjusted for non-responses when relevant (calibration weights). One must however bear in mind that the final product will not be a value-added indicator in the sense discussed above but an indicator on sales value or sales volumes when deflated.

The step by step estimation method sketched out above is deliberately made simple. This is done to illustrate the basic steps in the general estimation process. The discussions above have hopefully made clear that the selection of an estimator and the estimation process is a result of the whole process for constructing a survey.

Step 5.

For further aggregations combine the sector estimates with the set of value-added weights. If the survey results are to be published as indices, include that in the estimation process. As a reminder – for the index calculations a set of population estimates covering the base period must be available. Use the average sector values for the base year measured in base prices as the base.

All aggregates should be stored in the output database. A check of the aggregated results is strongly recommended – all aggregates should be graphed (automatically). Include the survey supervisor in this analysis.

Utilising the availability of additional information for another variable that is correlated with the variables observed, might enable the use of the rate estimator. This might be structural data on sales values, the number of staff employed in each unit etc. The number of staff employed among the non-sampled units in the stratum population is known from the Business Register.

The method to be preferred should be evaluated making comparisons between the monthly survey's annual results and the corresponding structural surveys when possible. The quality of any additional data – either weights from the structural surveys or the employment data from the Business Register – should also be part of the evaluation.

The preferred solution is built into the IT system and used in the estimation process.

The sample survey estimates and the corresponding population estimates are to be stored in the output database. The population results are to be used when producing tables for publishing.

3.4.3 Historical data – time series

Most STS statistics have a main focus on the time dimension but might also for some purposes be utilised for cross-sector analysis. Having a main focus on the time dimension also emphasises the importance of presenting historical data – time series to the user.

The INE has for some of the current STS statistics predecessors that might be linked to the new ones. This should be done as far as possible.

3.5 Analysis for quality

Analyses on special issues; structure of effects of non-response; changes in laws, in definitions, in adjustments made on the questionnaires etc – all expected to affect the quality of data received.

Analysing the output series is important for the total quality evaluation. All series are to be graphed – in the time dimension – to identify deviating performances at the sector level. This might be sector level performance untypical for the series and the month concerned, e.g. unexpected large changes from the preceding month or the corresponding month last year. An untypical sector level performance might be explained by additional information received from the province delegations or known to the Maputo Central otherwise. Additional information from a province might be of the type: one important unit has a stop in production due to vacation, stop for one or several days due to technical problems, strikes etc. If such types of information are not utilised it might be the cause of untypical sector performances.

During this analysis the supervisor must bear in mind any changes in the response rates. It is of utmost importance to separate impacts due to variations in the response rate from actual changes among the surveyed units. Variations in the response rates – without making corrections for this by imputing missing values – tend to produce "surprising" changes.

One should also be aware of any changes in the population (or among the sampled units) due to mergers or de-mergers of units. Such changes might create shocks in the sample by stratum estimates not necessarily reflecting the true population by stratum changes for the variable in question.

All suspected errors causing such changes are to be followed up – either by contact with the units concerned and/or by direct contact with the provincial delegations. Any new information should be included in the raw database (and the production database) or should be used for correcting the factual basis for producing the sector by stratum estimates. And – finally – the estimation process is to be repeated and all output databases updated.

When long time series for the aggregates are available for a survey, seasonal adjustment might be a relevant tool for identifying the so-called unexpected performances in series.

3.6 Meta-information – improving user accessibility to the statistics

Most users will from time to time need a documentation – or metadata – presenting how statistics are produced. Some users require detailed information for in-dept studies while others are confident with having a broader picture on how statistics are produced.

For internal purposes the INE must have a detailed documentation presenting all details concerning the principles and methods used in the production lines.

As a long-term target the INE must establish different documentation serving different purposes and needs. As a short-term target the INE should produce a presentation of each of the STS statistics following a standard model. It should have a maximum length of 2 - 3 pages. See annex E for an example.

Such declarations should be published in the *A Conjuntura Economica* – one at a time, and the declaration should also fit well into a system for an Internet dissemination of statistics and metadata. Using the Internet has several advantages in that the declaration will be made available along with the current results for all users having access to the web sites. Paper copies should also be made available to the users on request.

4 A new IT-system – a strategic part of the STS initiative

This chapter presents a draft plan for a new IT-system. The presentation is based on the chart shown below which outlines the main processes constituting the STS system of Mozambique.

The upper part of the chart represents the data capture process, which mainly is the responsibility of the provincial delegations located in the main cities in the province. The main activities – in a national statistical context – are support and service to the respondents, distribution and collection of questionnaires to the respondents, data entry and controls. When errors are identified the units concerned are re-contacted, and anticipated errors corrected or verified as being correct. The provincial process ends by mailing a file to the Maputo Central (MC) (in fact – sending the questionnaires). The MC is responsible for the processes described in the remaining parts of the flowchart.

The experience so far is, however, that the provincial processes do not function well and this might be one of several causes for large non-responses from the delegations. Emphasis must be put on the development of the province IT-systems (Data entry system) and especially improving the basis for an efficient flow of data to the MC. An e-mail system built into the provincial applications seems to be an extremely important function when aiming at an increased speed and timeliness in the production of the statistics.

The main processes of the MC production system – as described here – are mainly identical to what is found in other statistical institutions. A modern IT system does not exist in the MC today. Most statistics within the INE STS system are produced using fragmented IT systems – normally based on Microsoft products like Access, Excel etc. As for the province processes – neither the MC production processes seem to work efficiently – neither the manual parts nor the IT parts. The ambition is, however, to establish a modern production system based on databases for storage of raw data and production data as well as bases for storing time series for the aggregates resulting from the estimation process. The database system must furthermore provide a basis for an efficient and sound data administration. The INE has made a strategic decision on using Access for the IT production systems.

In the flowchart each of the major areas is given a number (e.g. (4.2)). This is a reference to the corresponding parts of this chapter.

4.1 Introduction – summary

A new IT system will have a strategic role when improving efficiency of the STS production lines. Vital for improving is that the data entry application to be used in the provinces is developed and implemented shortly. The application must be prepared for the Internet – and for sending data to the MC. Most likely this is the initiative – in the short run – that can speed up the flow of information from the data capture activities throughout the country – to the MC. To improve efficiency in the MC a new IT system seems to play an equally important role. This will be a system for receiving and storing provincial data (data administration), improve efficiency during data editing and estimation.

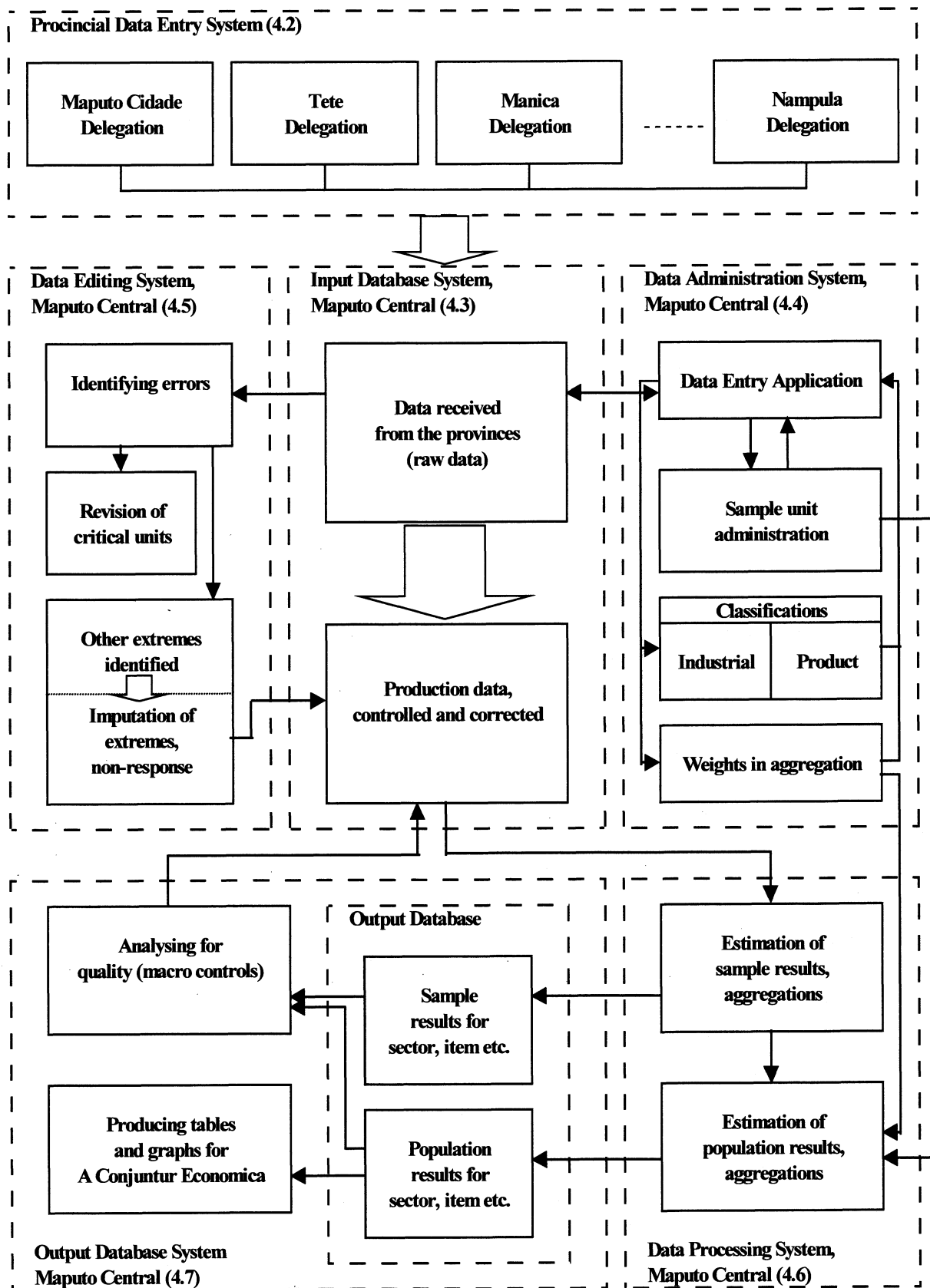
One should start establishing the system for one of the STS statistics. For a broader implementation the system established should be used as a model for all surveys. The main modules will be the same for each survey, but minor amendments in the database structures and programming codes will be needed. This is a cost-efficient way of developing IT systems.

Planning a new system in detail is a time-consuming process requiring a staff with high competency in the professional parts of the job. It is though recommended that the INE utilises the competence of short-term consultants for planning and assisting in the development of the more complex parts – especially the data control and estimation processes. The IT staff responsible for entering the programming codes for the system must be involved from the very beginning.

A new system will in several ways differ from the traditional more fragmented system. One main difference is that the new system will comprise most all processes of a typical production line. Some of the most important parts of establishing the system will be:

- ✓ Building a raw data database (RD) for storage of raw data received from the provinces and a production database (PD). The PD contains raw data – automatically loaded – and edited data. The latter is to provide the data for sample estimation and further aggregation to a national level.
- ✓ Building a data entry system to be used at the province offices. This will be an application for data entry and micro control of data. The data entry systems for the province offices should also be prepared for data-entering information about the establishment and changes. This is to be used by the INE Business Register as an important input for keeping the industrial population updated. E-mail should be used for sending all types of data to the MC.
- ✓ The MC data entry system will have additional functions enabling the unit to receive and load data to the RD. During loading – controls for doublets etc. will be performed. The application will also be used in administration of classifications used in the production and for updating weighting information used in aggregation.
- ✓ For an efficient administration of units participating in the survey concerned, the MC application should include a survey register routine. This is for the keeping and maintenance of all establishment information used during data capture (province, addresses, telephone no., fax no. etc.), and any other establishment information (industrial classification code, number of employees, annual sales) to be used in the production. This routine must communicate with the Business Register – flows of information going both ways.
- ✓ Develop a routine for error identification, check and imputations – based on the data stored in the RD. The routine will report errors (or extremes) to the operator according to pre-defined criteria. This part of the application will be based on several test methods that have proved efficient for this purpose.
- ✓ Building an output database comprising time series for all relevant population aggregates. Each series stored must have a name (for identification) allowing for an easy access to output series. The INE is recommended to develop some standard principles for names of series in the databases.
- ✓ Develop routines for producing estimates on sample level being a part of the basis for estimation of the population aggregates. The survey level could be on industry group level (sector) or on a more detailed level (e.g. item or any other groups of data considered as relevant). The survey estimates (sample and population) could be totals or averages of the variables observed and stored in the PD.
- ✓ A routine for macro analysis and production of standard tables used in publishing. The standard tables should be prepared for publishing on paper as well as for dissemination using the Internet.
- ✓ Operating a production line system requires a set of catalogues for keeping the classifications of relevance. This could be the standard of industrial classification (ISIC); the item code classification used (CPC) etc.

The STS production system



As an increasing number of provinces will be linked to the MC – consequently a flow of information will be received on a monthly basis. The data to be observed (values and volumes) will be an important part of this. The need for the future will, however, be to receive, store and systematise all types of information from each province reflecting the changes in the sampled units (shut-down, changes in address, some

items not produced – changes in item descriptions, changes in quality on the products observed etc.) Any changes in the establishment information must be passed on to the Business Register.

The remainder of this document sketches out a draft plan for the IT work to be done. In the presentation of the main activities each subject is briefly presented as an introduction and objectives are given. In addition related activities are presented and discussed. Each presentation ends by presenting a list of important outputs.

4.2 Provincial Data Entry System

This is the data entry system to be used in data entering – values, volumes and units etc. The application will primarily be used in the province delegations. The design of the system should aim at securing the flow of information into the Maputo Central – a flow containing information on changes of the observation unit (establishment), changes of items – new items, deviations from the predefined descriptions, interviewers quality judgements etc. Flows of data will be going both from the province to the MC as well as the other way – e.g. provincial statistics sent to the provinces (that are consistent with the national aggregates).

Objectives:

The overall objective is to establish a new system for data entering improving the local control of data, and to improve the insight of the MC into current coverage of units and breakdown used in observation – and recent changes observed.

Activities:

- Establish new routines for entering observed value and volume information – and units. The new routine should provide information on changes in the establishment and the products / breakdowns used in each of the surveys. This type of information will be needed in the MC as a basis making sound judgements when handling changes.
- When changes are made in the establishment information – address, telephone, changes in industrial activity etc. – a copy of the establishment record (extract from the database) should automatically be mailed to the persons responsible for the Business register.
- Routine for entering new items and descriptions of products / breakdowns, and new establishments will be needed. This is a very important part of the new entering system.
- Routine for extracting data to be mailed to the Maputo Central should be developed. Testing of this routine should be made on pilot versions to catch possible problems in an early phase.
- To decide how data are to be received by the Maputo Central and how the various types of data in an efficient way can be routed to the relevant tables in the base system. An automatic updating procedure based on the files received should be considered.
- One should consider making an automatic report to the operator for the current month (screen / paper) showing what has been done – number of data elements entered, number of changes, product descriptions that are changed, number of new establishments etc.
- When the database system and a pilot of the data entry system are ready – to enter the types of data to be used and start testing the entry system. In light of this – make adjustments whenever needed.
- Training of the staff in using the new system
- Documentation is very important – both technical (IT) and user-related

Outputs:

- A new data entry routine
- Automatic routine for routing data to the relevant tables in the Maputo Central
- Routine for automatically sending of a mail containing changes made for one of the sampled units – to the Business register.
- Documentation of the technical aspects (IT) and the user-related parts

4.3 Input Database System

This base system will comprise all value and volume data – on establishment level – received from the provinces as well as additional information on any types of changes etc. The system will store the data received from the provinces in an input database – as they were delivered from the province. Furthermore – a production database for storing the value and volume data used in the computations should be developed. The two bases will differ in that the production database will contain any centrally collected information and cells that have been imputed due to non-response or due to statistical purposes.

The main reason for establishing two databases is manifold but two should be mentioned. It is important to separate the results from the data collection phase i.e. in the provinces – from what has been done in the Maputo Central. Furthermore the separation allows for making analysis of the effects of the MC work and its impact on the overall estimations.

Objectives:

The overall objective is to establish a database system, which is capable of handling a large flow of data from the provinces to Maputo. It is important that the new base system is prepared for handling all provinces.

Activities:

- To decide how the database structures shall be – variables etc.
- To decide how the various types of information received from the provinces are to be stored to secure an efficient use of such information during the current revision work.
- To decide which item / breakdown code system that should be used. The code system should be in line with international recommendations for items (CPC / CPA). All standards and classifications are stored in catalogues. Item or breakdown codes not being a part of the catalogues are not to be entered into the data entry system (verified through a check towards the catalogue concerned). The strict restrictions on this matter are made to remove data entry errors and to avoid the existence of several codes for one item.
- Implementation of an item code / breakdown system
- Creating an input database containing the observed values, volumes and units from the provinces as well as any other information concerning changes in products or the units (establishment) that will be received.
- Creating a production database containing the data to be used in the computations of the provincial as well as the national aggregates.
- To decide how the database structures shall be – variables, industrial codes, product codes etc.
- A routine for loading data from the raw database to the production database. To cope with a situation where normally data from the provinces will be received in several rounds (for one unit), the loading routine must be able to work selectively e.g. allowing for changes in the previously (for same month) reported data. Bear in mind that adjustments in data received from the province data are only allowed when operating in the production database.
- One has to decide how the data outflows and data inflows to the two databases shall be. The production database will provide other parts of the production system with the relevant information e.g. to the micro index computations.

Outputs:

- A database containing province raw data (RD).
- A production database (PD).
- A new item code system – implementation.
- Automatic loading of data from RD to PD.
- Training of the staff
- Documentation

4.4 Data Administration System

This is a system comprising functions like data entry, administration of sampled units, administration of classification catalogues/tables and the weights used in aggregation (when relevant).

4.4.1 Administration of data – receiving information from the provinces

This application will be used for maintaining information about the sampled units of each survey i.e. the name of the establishment, the enterprise, the address, the name of contacts in the firms, telephone, keeping the industrial code, keeping record of the response on a monthly basis. In fact this is a small-scale business register mostly based on information from the national business register. In many countries such registers are part of a national business register – as a sub register.

The survey register will have as a main purpose to keep record of changes in the unit information as well as tracking the unit responses on a monthly basis. This type of information is important for the maintenance of the population of units in the sample. All types of verified changes must of course be reported to the national register securing a high quality of the population of units.

This application will also be used for receiving and loading data from the province delegations. Automatic loading to the relevant databases should be considered based on a standard reporting format to be used in the province delegations.

From this application one should be able to make changes in standards and classifications – industrial codes as well as product codes. Any information indicating deviations from the predefined descriptions, interviewers' quality judgements etc. should be stored in these bases. A great deal of the information on changes is to be received from the provinces. When the Maputo central makes changes in the samples i.e. including new units or excluding units, this type of information will flow back to the province delegations.

Objectives:

The overall objective is to establish an application for keeping track of the sampled units based on the national business register in the first place and the current flow of information received from the province delegations concerning changes in the units.

Activities:

- Establish the data base structure needed – could in large be the same as in the national business register. What has to be added is a function for keeping track of the monthly response merits of the sampled units.
- Build the application for making changes in the establishment information – address, telephone, changes in industrial activity etc.
- Build the application for receiving and loading data to be received from the delegation offices.
- Develop a routine that copies an establishment record (extract from the survey database) to be sent to the persons responsible for the national business register.
- A routine for an automatic update of one record – or groups of records – from the national business register must be developed.
- Routine for entering new items and descriptions of products / breakdowns, and new establishments will be needed. This is a very important part of the new entering system.
- Routine for extracting data to be mailed to the delegation offices should be developed. This routine has to be standardised and used in the delegation applications.
- Establish a system for an efficient update of the tables in the databases in the Maputo Central. The various types of information must in an efficient way be routed to the relevant tables in the base system. An automatic updating procedure based on the files received should be considered.
- One should consider making an automatic report to the operator for the current month (screen / paper) showing what has been done – number of data elements entered, number of changes, product description that are changed, number of new establishments etc.
- Training of the staff in using the new system
- Documentation is very important – both technical (IT) and user-related.

Outputs:

- A new database comprising unit information for the sample.
- A routine for the administration of data received from the provinces.
- A data entry routine for mending the unit information as well as for entering of unit information sent directly to the Maputo Central.
- Automatic routine for routing data to the relevant tables in the Maputo Central
- Routine for automatically sending to and receiving from the national business register.
- Documentation of the technical aspects (IT) and the user-related parts

4.4.2 The data entry system

This will to a large extent be equal to the province systems, but requires functions that will be specific for the Central Unit. The Central will have to make changes in the catalogues comprising standard and classifications and the weight information.

Objectives:

Improve the efficiency of the revision work in the Maputo Central.

Activities:

- The new data entry system must provide support for the use the Maputo Central will have when entering centrally collected information, making special adjustment in data due to changes in units etc. or for statistical purposes.
- The entry system should also be prepared for making changes in the classifications used as well as the weighting structures.

Outputs:

- A data entry system adapted for the Maputo Central's needs.

4.4.3 Standards and classifications

The standards and classifications used during estimations should be stored as a part of the IT system – in catalogues or suitable tables (in a database system). The catalogue should also include the text or description of each sector and the aggregated levels. The catalogue should also be prepared for two types of texts: a long text equal to the standard formulations, and a short text more suitable for use in publishing. If publishing in English should be relevant, the catalogue should also contain this text. Texts must be the same for all provinces and at a national level – to enable comparisons across the country. If the text used in publishing is different for the provincial and the national results, the short text should be developed for this purpose. It is not recommended to use different texts for the same type of activities.

The major advantage of having the classification (or several classifications) in a catalogue is that this simplifies any type of future updates of the classification to be used.

Objectives:

Improve efficiency in the production routines for each province and at the national level.

Activities:

- To consider which classifications that should be used.
- Produce a table linking each type of items / breakdowns to its corresponding 3-digit ISIC classification level or – if relevant – use the national version of the ISIC. Include the long (but precise) text reflecting the text used in the international classification being a model. Include a short text if needed for publishing. This link-table should then be loaded into the STS IT system. The data entry system discussed above should be prepared for allowing changes in the classifications to be made.
- Training of staff
- Documentation

Outputs:

- A complete catalogue to be used in the system for estimating aggregated CPI on provincial and national level.

4.4.4 The weights for aggregation

As discussed in the preceding parts of this chapter, the weighting data should be stored in a catalogue including both item / breakdown level weights and weights for the aggregated levels (3-digit, 2-digit, 1-digit) for each province. Each type of item should be specified with the item code and the corresponding weight. When using the rate estimator the weights (or rates) are to be estimated each month and stored in the catalogue.

Having the weighting structures of each province and the national weights stored in one place will for future revision substantially simplify the updates of the weights.

Objectives:

Simplify future works on updating the weight structures and to enable aggregations of the provincial and national results.

Activities:

- Make a table linking each item level code to the corresponding weight – for each province. The data entry system discussed above should be prepared for allowing changes in the weighting structures to be made.
- Establish the weights for national aggregates – to be included in the weighting catalogue.
- Implementing the catalogue into the business statistics system.
- Training of staff
- Documentation

Outputs:

- A complete catalogue to be used in the system for estimating aggregates on provincial and national level.

4.5 Data Editing System**4.5.1 Identifying errors – micro controls**

This is an important part of the new IT-system and the implementation of new methods to be used during revision. The experiences from a vast number of countries have shown that too much time is spent on making personal judgements in this part of the revision. Misjudgements and lack of symmetric information have led to an "over revision" in many situations. The large number of provinces and the number of establishments and items / breakdowns will require a need for an efficient system for control and error identification.

This part of the revision focuses on several aspects of the sampled units. The common approach is to study the recent changes – the variation over time – in values, volumes, unit prices (value/volume – only relevant for some of surveys), values / number of employees, (value – wages)/value etc. For statistics producing sector aggregates (e.g. 2- or 3-digit level according to ISIC) it is also recommended to study the variation in space i.e. across the establishments covered by the aggregate concerned. The examples below illustrate the task of checking data – over time and in space. There should be no reason to focus only on the establishment performance over time. Several types of checks should be built into the production routines.

The first type of checks focuses on the establishment and the performance of the item / breakdown for each of the data reported – over time. This is the basic type of checks done – mainly focusing on the variation in the unit prices estimated from production values and produced quantities.

The second type of checks emphasises the establishment performance in space i.e. compared with other establishments in the same sector. This type of check should normally focus on one or several indicators giving a good description of the relation between the establishments. One should consider the use of an indicator like "(sales value – wages) / sales value", which gives a rough picture of the net sales value (though not corrected for the cost of any inputs used in production). Another indicator could be "production value / number of employees", which should provide an estimate on the production capacity of the average worker.

These indicators will of course vary between the establishments on sector level. This is simply due to differences in the establishment production capacities (and variations in the current utilisation rate), due to differences in price levels and variation in these (across establishments) etc. In this context we do not focus on the performance of each establishments (over time) – although it could be done. We are instead working on the sector performance and the variation of this – measured through the average – variance (standard deviation) relations or the median – quartile relations. These statistical parameters describe the statistical distribution of the sector performance. If the indicators are stable or changes slowly over time (the response rate should be on an acceptable level e.g. at least 75 per cent), we should be able to identify establishments changing their performance in space compared with the average performance of the sector. The establishments that are flagged as extremes should be treated manually by a subject matter specialist to identify possible errors in the data. We should also be able to identify changes in the sector performance over time, which might be due to variation in the non-responses within the sector. E.g. establishments responding last month were missing this month. There are, however, all reasons to believe that the sector performance varies somewhat over time. Such variation should be used to establish acceptance levels for the sector performance.

It must be added that the examples used in the preceding paragraphs are not to be adopted and put into operation – without analysing the qualities of the indicators i.e. their ability to function as intended. The main idea is to describe the main features of such a control system, possibly outcomes and the use of these. Checks should be based on several indicators that are complementary.

If one is publishing at the province level, such checks should also be considered for each set of provincial data. Establishments or items / breakdowns identified as extreme are to be revised, while the other non-flagged items are to be considered as accepted and verified – and no further action is needed.

Objectives:

To reduce the corrections made during the current revision phase to an efficient minimum, and to avoid unnecessary revisions of the data.

Activities:

- Routine for extracting the micro data from within item / breakdown level from the production database – for each province. Establishment code, item / breakdown code and observed data for the current and previous period are needed.
- Programming the system of mathematics for the control routine
- A window for fixing the parameters to be used in operating and starting the system.
- A report showing items, breakdowns or establishments for further control should be generated by the system.
- Routine for making flagged units available for the operator according to a sorted top-down principle – loaded to a separate table ("table of outliers") – to be treated in the data entry system. All information about changes reported for the flagged items in the province data sets should be made available.
- Training of staff
- Documentation

Outputs:

- A full working routine for identifying errors in relative prices using all information available

4.5.2 Treating extremes – imputations

This is another new element that should be introduced into the business statistics IT system. Studies into the field of imputation have shown that the subjective element by the operators or the person responsible for revision varies over time. Changes in staff in this editing area will also contribute to humanly caused variations. The reason for introducing this part is to secure an over time similar treatment of outliers by using the same methods from month to month.

The methods available for the types of controls mentioned focus on the extremes to be found in the tails of the statistical distributions in question. Another type of problem relates to the inliers i.e. responses from establishments that – over time – tend to show no change or show changing patterns that differ substantially from the typical performance of the establishments in a sector / stratum. The inlier occurs in various contexts, but will be most apparent in sectors where all establishments are fairly homogenous – as concerns the produce and markets. A deviating changing pattern for the inlier does not necessarily mean that the value reported is erroneous. In a sample survey context where the number of observations in a stratum / sector might be scarce, it is recommended that the inlier is subject to further examination and follow-up towards the respondent.

In every period – outliers are expected to occur, which can not be treated by automatic procedures. In many situations the number of outliers that can be treated manually has to be fixed according to the capacity of the staff occupied with such tasks on a regular basis. E.g. if the staff capacity during the limited number of days available for revision should be estimated to 25 cases, then the 25 most extreme outliers loaded to the “table of outliers” (see (4)-(5) above) are treated manually by the staff. If more than e.g. 25 items are flagged, the number of units above 25 is handled using the automatic routine.

In automatic imputations a number of rules could be used. It is recommended to use a fairly simple set of formulas in such routines. E.g. using hot-deck variants like the average or median increase of the non-flagged items. This could be a system comprising several imputation methods, which are automatically called for – depending on the characteristics of the extremes identified. A simple method could e.g. a unit non-response occurs in May. For a unit normally having stable values and volumes over the calendar year (this could be the result of an analysis on the historical time series for the unit). In this case a copy of last month's data could be reasonable. If not stable throughout the year, another method could be automatically used.

Objectives:

To build in an over time consistency in the imputations for the provincial and national aggregates, and to improve the efficiency of the manual adjustments (subject matter specialists work with the problem units – and not with minor or irrelevant ones).

Activities:

- Routine for extracting the rest of the outliers from the table of outliers to be treated automatically – for each province.
- Programming the system of mathematics for the imputation routine.
- A routine for rearranging which units to be treated automatically and manually.
- A report – showing the units that have been subject to automatic imputations.
- A routine for updating the production database with the unit subject for automatic imputation.
- Training of staff
- Documentation

Outputs:

- A full working routine for automatic imputations

4.6 Data processing – population estimates – aggregations

The aggregations will be based on the item / breakdown level results produced in combination with the weighting catalogue and consumption classification catalogue. The target for this operation is to produce sector level data.

Several options are possible when estimating the population aggregates – for provinces, sectors or combinations of these. Two common ways of handling these are:

- The rate estimator is often used when the variable observed (e.g. production value) is correlated with another known variable e.g. number of units or number of persons employed in the establishment. The method might contribute to correct for non-responses by allowing for variation in the rate from time to time according to the number of units (or number of employees) that are responding and which that are not. The method allows for some variation in the response rate of the survey concerned – when having a neutral non-response. It will not be possible to compensate or reduce the effects of biases resulting from non-responses. Estimating within strata having a group of homogeneous units (same size or not too large variation in size) might be efficient.
- Using fixed weights e.g. value-added from the annual national accounts or the annual survey for the same sector when estimating the production index totals. Even this method will be highly dependent on good quality in the sample survey estimates, and that care has been taken of possible bias resulting from non-responses.

The aggregates should be estimated – at the lowest – the 3-digit or 2-digit level. This will vary across the surveys. For some the aggregates will be based on sums of the value figures, while the use of indices might be more relevant in other surveys. Most users know the traditional fixed base Laspeyres formula and this should be considered when publishing indices.

As an important part of the aggregation system a number graphs and tables on the main aggregate levels should be produced as a standard. The information should be used for macro controls on the data / time series for each aggregate level.

Objectives:

Estimate the sector aggregates for each province and at national level.

Activities:

- Routine for extracting the item level data, weights or producing the rate to be used in aggregation and ISIC classification codes – for each province from the respective catalogues or bases.
- Establish the weights or rates for national aggregates – to be included in the weighting catalogue when fixed weights are used. Or – estimate the rates to be used in monthly / quarterly aggregation (requires information on units responded and units not responding).
- Programming the system for aggregation
- A routine for updating the SA database (see (10)) – aggregated results.
- Training of staff
- Documentation

Outputs:

- A complete system for estimating aggregated CPI on provincial and national level.

4.7 Output Database System

Two issues are central in this part – the output databases and requirements, and the macro control part including publishing.

4.7.1 Output databases

The output database will contain two parts or bases – one for item or breakdown level aggregates (BA) and one for the sector aggregates (SA). The data should in both bases be stored as time series and it should be an easy task to extract full or parts of specific time series. The BA database will along with the classifications used in estimation and the weighting structures – constitute the basis for estimating the sector results.

Indices will be estimated for each of the provinces and at the national level, where series could be produced and stored in the database on 3-digit, 2-digit, and 1-digit ISIC and as totals across sectors – if relevant. What is published is a separate matter that should be made independent of what is stored in the database. For the SA one has to develop a code system for identifying the different series.

Objectives:

The overall objective is to establish a data base system, which enables efficient flows of data and storage of the aggregates produced. Easy access to the stored time series is required.

Activities:

- To decide how the data base structures shall be – variables etc.
- To decide how the code system for time series in the SA database shall be
- Implementation of the new code system for time series.
- Creating a BA database containing the estimated item and breakdown information produced for each province.
- Creating a SA database containing the aggregated series estimated for the provinces and the nation as a whole.
- One has to decide how the data outflows and data inflows for especially the SA database shall be. The production system for standard tables on paper and on the web-site (see (14) below) will be based on these data.
- Training of the staff
- Documentation

Outputs:

- An BA database containing provincial series as well as national series
- A SA database
- A new code system for time series – implementation

4.7.2 Analysing output – macro controls

A system part for producing the standard tables used in publishing on paper or on the Internet is recommended. This will improve the efficiency of the publishing part of the work and will most likely be a fairly simple task to put into operation.

4.7.3 Standard tables for publishing

A system part for producing the standard tables used in publishing on paper or on Internet is recommended. This will improve the efficiency in the publishing part of the work and will most likely be a fairly simple task to put into operation.

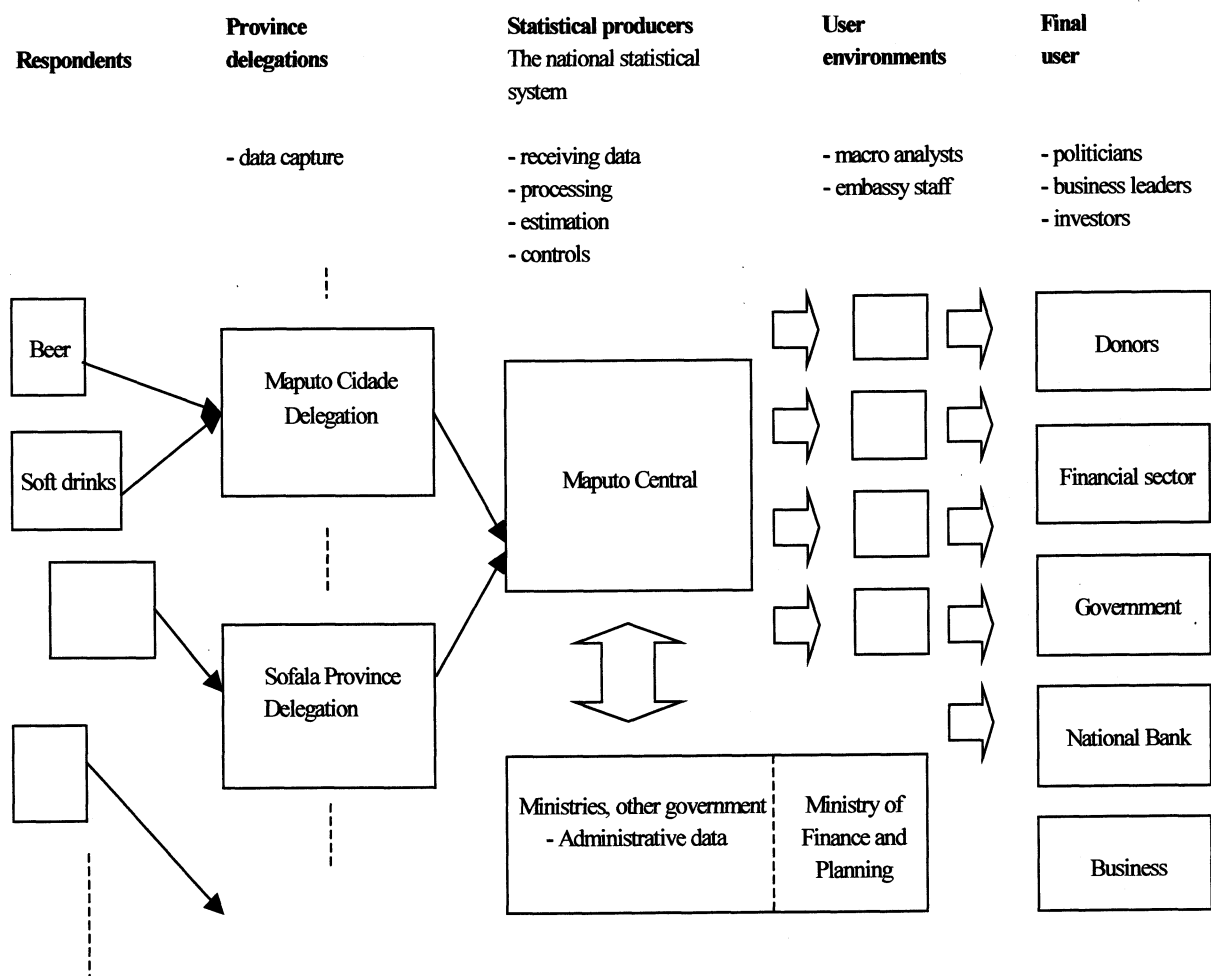
There should be no need for going further into this routine and I leave further planning on this part to the staff concerned. Developing such a system will give the users competence in operating and developing the system as a whole.

5 Dissemination with a special focus on A Conjuntura Económica

The user requirement for timeliness and punctuality – far from being fulfilled – calls for an INE initiative on providing advance information of the publishing plans. The first part of the chapter discusses the statistical release calendar as a tool for providing advance information concerning publishing plans to the users. The calendar is to be produced on a regular basis. The second part of the chapter has the focus on A Conjuntura Económica – the main channel for disseminating short-term statistics. For more information concerning dissemination policies in general – see Johansson and Åström (2000).

The INE's role in the information industry is normally being the intermediate producer – in the line of users / analysts systemising and presenting information on the business cycle, inflation etc. for the final use of the decision-makers. Due to this there exists a close relationship between the INE and the users and the requests for timeliness and punctuality from their customers.

The Information Industry



5.1 The statistical calendar – advance information on publishing plans

Most users want up-to-date figures that are published frequently and on time at pre-established dates.

- ☞ **The INE is recommended to produce a statistical calendar – updated on a regular basis (monthly) – and made available to the users both on paper and on the Internet. The calendar must contain pre-established dates for the publication of all statistics.**

In appendix D some ideas concerning the statistical calendar and contents are discussed. The appendix calendar provides an example that illustrates what is required for establishing such a calendar – in a technical sense. Some pitfalls to avoid are also briefly discussed.

For the dissemination of the statistical calendar the INE should utilise all channels available. To distribute the calendar using the Internet is very efficient. In a mainly paper-based regime – as today – this type of information should be made available on a regular basis as an attachment (or direct part) of A Conjuntura Económica – and in all other documents released by the INE. In most European countries the calendar is also disseminated – electronically – to media (papers, radio etc.), which distribute the information to their readers or listeners.

5.2 A Conjuntura Económica

A Conjuntura Económica (further – Conjuntura) was started about two years ago with the objective of being a channel for distributing STS. The idea does not seem to have been that of producing a macro economic analysis.

The number of copies (approximation of the number of users) produced every quarter is close to 110 including free copies distributed to governmental sectors (35) and media (newspapers, television and radio). The sales vary from quarter to quarter but normally some copies remain unsold. There are no complete sales registers which allow for identification of the non-governmental users. Some few have been identified indicating that this user group mainly comprises international organisations (IMF, World Bank etc) and national institutions (within the financial sector), and users within the donor community.

5.2.1 From a user point of view

Meetings were arranged with governmental and private sector institutions – and persons anticipated being readers of the publication. A general impression is that some few of the anticipated readers had seen the publication lately. Only one person seemed to be an active reader – primarily using A Conjuntura for gathering information for analysis of the economic situation in the country. Although not having a success when meeting all the anticipated users, interesting comments were given – mainly based on the experience in being a user of the INE statistics and on reading the INE publications in general.

Some conclusions from the meetings are:

- ✓ A Conjuntura does not seem to have a clear profile as concerns the target group or groups. Although the contents indicate an ambition to provide a basis for macro economic analysis – the speed in the production and dissemination of the publication do not meet the user needs. Financial sector information from the Bank of Mozambique and CPI data that are presented in the publication, have for months been available to the users. The publication seems in fact to be a reference publication gathering and distributing data that already are made available to the users.
- ✓ The presentations based on tables and graphs are in general expressing the same, and for some areas the main points in the information presented are hard to find. The presentation of CPI data works better in that the graphs based on longer time series give a reference for the users' understanding of the more recent trends in data.
- ✓ The lack of comments in writing on tables and graphs in most presentations makes it hard or almost impossible for an untrained user to find the interesting results. The competence of a macro analytical expert seems to be needed. The presentation of each sector must have comments that work supplementary and assist the user in understanding the results. When large and surprising changes occur it is recommended to comment on this by explaining what has happened. Some users recommend that A Conjuntura – on a quarterly basis – should include one page summing up the broad picture for the economy as a whole.
- ✓ Especially the financial sector information requires more background information in writing. There are e.g. several interest rates that are presented, however, without guiding the reader in the role or

functions of the different series. Some comments defining the variables presented could be given in a footnote or separately in a textbox for such purposes. The use of footnotes or textboxes is in general recommended for many of the tables.

- ✓ Some of the tables contain results that are considered as surprising in the larger context. Many users have long experience in finding errors or what are suspected errors.
- ✓ Most users seem to have a preference for sector information instead of item information, which is dominating the coverage of most sectors. The tables should primarily present value information – in Meticaís or USD. One currency is recommended used throughout the publication. A more complete presentation should include volume indicators as well (values in fixed prices or volume indices), and – over time – corresponding price indices. Table contents should also be developed to include annual information as well.

In addition there were some comments on the lack of reference to the source of some of the tables and explanations to the codes used in the tables.

Based on the meetings and my own evaluation of the publication, changes are recommended. A Conjuntura does not seem to be a channel for reaching the professional user. In general most data have been available to the users months ahead of the release of the quarterly publication. Many of the tables are complicated and require supporting text to enable an easy access to – or understanding of the results. These comments indicate that A Conjuntura in fact does not reach the target group of readers – as planned. One option is to reconsider the target group and design the publication and contents especially for that targeted group of readers. Another alternative would be to change the frequency of A Conjuntura – to be released on a monthly basis!

Whether changing the frequency or not, some changes have to be made as concern the contents and layout of A Conjuntura, and several changes recommended are specified later in this chapter. But before entering into the details some consequences of having a monthly publication should be discussed. The recommended changes in contents and layout should be independent of changes in frequency.

5.2.2 Statistics on a monthly basis

☞ **The INE is recommended to change to a monthly frequency of A Conjuntura.**

All separate publishing of sector statistics within the STS including the CPI are discontinued and released through the new monthly A Conjuntura.

The date for releasing the publication is co-ordinated with the dates for releasing the CPI – to avoid delaying the CPI release.

The contents and size of the publication should be allowed to vary from month to month. This makes it possible to have a special focus on what is new in each monthly edition i.e. on the statistics recently produced and to be published. To create continuity and to provide the users with an overall picture of the economic situation, each monthly edition should include a table summarising the recent figures for all business sectors. This table should only contain the sector aggregates and information presenting growth rates etc.

All short-term statistics will be released in one publication. This will widen the existing target group to include all types of users – also the professionals. The publishing plan for each of the STS statistics should be co-ordinated with the release calendar for A Conjuntura, which allows for having different speed in the production plans for the sector statistics.

The STS initiative – and the need for disseminating the statistics at fixed dates each month – will lead towards the need for discussing the publishing frequency of A Conjuntura. The alternative for the STS

statistics is to publish separate reports for each of the statistics – or press releases – to avoid delaying the speed in publishing.

Production plans for each of the STS statistics should be co-ordinated with the time schedule for releasing A Conjuntura.

A monthly publication requires an efficient system of production and dissemination. The productions' technical aspects are discussed in a separate report on printing and dissemination – Johansson and Åström (2000) – and will not be discussed in this context. A monthly A Conjuntura will most likely require that the editorial part is administrated by a person – not directly involved in the production line for the statistics.

5.2.3 Improving the user value of the publication

This chapter presents the recommendations on contents, layouts, tables and graphs. The initial part focuses on ways to increase user value through editorial changes aimed at making a whole out of the elements in the report – in a consistent way. The use of text is central in this part. In the last part of this chapter the focus is on the details.

The recommendations given are not of the urgent type. Change in A Conjuntura is recommended put into operation from next year on, when STS publishing is expected to start. It is, however, a need for initiating the planning process already this year.

Contents and layout

- ☞ **The front page should be standardised using a symbolic picture. This might be as a background for the presentation of the contents of the monthly publication. The INE logo should also be a part of the front. Only the month, year and number information should be changed.**
- ☞ **The second page should be used for a brief presentation of the INE – like what is found on the last page in the current A Conjuntura. The second page should also include a small table explaining the codes used in table cells when data are missing (na), a value is equal to zero (-) etc. The second page could also include brief presentations of the statistics e.g. annual or periodic surveys that are released.**
- ☞ **The third page should contain a table giving the last data available for all STS statistics, CPI and others released in A Conjuntura. A brief text summarising the current situation should also be included. Together – maximum one page.**

Every quarter (the February issue, May issue etc) the table should be replaced by a somewhat larger summary based on all STS statistics and other sources available. The text should be maximum one page. In an initial phase the INE will need some assistance in the analysis and to summarise the analyses. The idea has been discussed with Gabinete de Estudos, who supports the initiative. Members of the staff will – if considered of interest for the INE – assist in the analysis, writing and training of the INE staff.

- ☞ **The last page should have a permanent table presenting the key indicators (a selection has to be made) from the last annual data from the national accounts (NA). The table must immediately be updated when new results are available from the NA. This table functions as a reference and provides a consistent frame for interpreting the STS statistics that are currently published during the year.**

If considered relevant – some few comparable indicators for neighbouring countries (from the Live DataBase) could be included in the last part of the table.

- ☞ **During this mission a documentation of the CPI has been produced – see appendix E. To improve the user accessibility to basic principles and methods used in the statistics – in general – such brief presentations should be produced for all statistics and published in A Conjuntura.**

The INE is recommended to include – one at a time – documentation (3-4 pages) in A Conjuntura, which is the main channel to reach the users. The documentation should also be made available on the Internet.

This concludes the recommendation as concerns contents and layout (the profile). These initiatives will improve the user value of A Conjuntura.

Although this might – at a first glance – seem complicated one should bear in mind that large parts of the four pages discussed remain fixed during the year. To make this functioning on a regular basis requires – first of all – clarity in the production routines and a motivated person – good at writing and having the capability and interest for analysing the statistics.

Recommended changes on details in contents

The tables and the graphs should in layout be standardised (as much as possible) – for each sector providing the same types of information allowing for relevant comparisons across sectors and over time. The use of footnotes is stressed for clarifying what is estimated, to define complex variables presented etc. and the use of references to the source. These issues are also of relevance when publishing graphs.

- ☞ **The STS should publish comparable data across sectors. For this purpose tables based on production values or sales values should be published for all sectors – in current MT prices. It is also recommended to use the same scale for all sectors tables, i.e. measured in 1000 MT. See the example below.**

TAB. new1 - INDUSTRIAL PRODUCTION, VALUE BY SELECTED INDUSTRY GROUPS. Current 1000 MT

Industry	1998	1999	98-99	1999				2000				VH
	Ano	Ano		I Trim	II Trim	III Trim	IV Trim	I Trim	II Trim	III Trim	IV Trim	
Manufacturing, total	244245	396644	62,4	60022	57403	56982	69838	68989	63410	112012	152233	118,0
Food industry	114625	123562	7,8	26394	29416	29646	29169	31413	28927	31913	31309	7,3
Clothing	4208	4542	7,9	1347	867	874	1120	1241	978	1178	1145	2,2
Wood and wood products	69555	90807	30,6	15730	16910	13975	22940	19315	20687	22552	28253	23,2
Chemicals and chemical products	53673	58141	8,3	16188	9723	11821	15941	16810	12648	11514	17169	7,7
Plastic and mineral industry	2184	747	-65,8	363	487	666	668	210	170	160	207	-69,0
Aluminium	0	118845	-	-	-	-	-	-	-	44695	74150	-
Domestic trade, total	203786	213188	4,6	48472	44928	50531	59855	49484	50045	47229	66430	11,0
Retail trade	187928	196271	4,4	44241	41366	47210	55111	44897	46483	43562	61329	11,3
Wholesale trade	15858	16917	6,7	4231	3562	3321	4744	4587	3562	3667	5101	7,5
Cars and petroleum products	10946	139912	1178,2	1692	3320	2573	3361	31742	35220	29297	43653	1198,8

Fonte: INE - Inquerito mensal as empresas de producao Industrial

- ☞ **Following what has been discussed previously in this report, all tables will contain population estimates. For tables where other estimates provide the basis it is strongly recommended to clarify –briefly – what is actually presented (estimated). For this purpose – use a footnote to the table. The footnote should also be used for defining complex variables.**
- ☞ **Each table should always have a reference to the source. The source for statistics produced in the INE should always refer to the INE – possibly in addition give the name of the survey as supplementary information. The reason for using the INE as a reference is to make clear for the user which institution that is responsible for the statistics. The reference to the INE should also be used when estimations by the INE are based on data from other institutions.**

Appendix A The terms of reference

A short term mission to the National Statistical Institute (INE), Maputo, Mozambique

Topic

Short term statistics – methodology and dissemination

Background

INE has recently launched revised monthly surveys for industry, construction, transport, commerce, hotels and restaurants and services. The surveys are still in the starting phase, and several adjustments of the methodology applied are likely to be necessary during the first year. Responsible for the short-term surveys is DESE (The Directorate for Sector and Enterprise Statistics). DESE is divided into two departments, DEBA, which is in charge of the industry and construction surveys, and DESSIC, responsible for the commerce, hotels, restaurants, transport and service surveys.

The short-term statistics are published on a quarterly basis in the publication “A Conjuntura Económica”. The new monthly surveys opens for alternative ways to publish the results. “A Conjuntura Económica” is now in its second year, and having the new surveys in mind, it would be useful with an evaluation at this stage. The responsibility for “A Conjuntura” is placed in a department responsible for Consumer price statistics and Short term statistics (DPE), which is part of the DCNIG (Directorate for National Accounts and Global Indicators).

The Task of the Consultant

- ✓ The consultant shall assist the Institute in formulating a methodology for short-term statistics production and dissemination. The following issues and items should be included in the work:
- ✓ Suggest methods for preparing the data collected in the monthly economic statistics surveys, among these methods for treating non-responses, methods for estimating totals, seasonal adjustments, methods for aggregated indices etc.
- ✓ Suggest ways to publish results from the monthly economic surveys (indices, tables, graphs etc)
- ✓ Do an evaluation of the quarterly publication of short-term statistics (A Conjuntura Económica).
- ✓ Suggest alternative ways of presenting results in the same publication to improve the dissemination of short-term data.
- ✓ Suggest new ways of dissemination the short term statistical information

Method

- ✓ The consultant will be working together with personnel in charge of production and dissemination of the economic short term statistics
- ✓ The consultant will meet with some major users of short term statistics
- ✓ The consultant should visit some institutions rendering important sources of information
- ✓ The consultant should present a draft report with major findings and recommendations by the last day of the period

The consultancy should be carried out during 3 weeks, starting 20 June 2000.

Appendix B The STS surveys of the INE – an overview

All surveys – excluding the external trade statistics which is based on administrative information from the Custom Authority – use the same sampling plan including all large establishments (> 50 employees) and a sample of medium sized and smaller units. The data capture work is done by the provincial delegations while the remainder of the editing process is done in the Maputo Central Unit. All surveys cover employment (number of employers – men, women and total) and wages and salaries. The measured value data excludes VAT and all other taxes levied on the products.

The presentation of the surveys does not comprise the CPI, but focuses on the new or recently revised sector statistics.

The manufacturing production survey

Collected data:

Production data on item level – quantities and values (both produced and sold) – monthly. Cover the relevant manufacturing sectors.

Actual publishing:

Item statistics (physical quantities) for a small sample of large establishments. Preliminary versions of the manufacturing production index and the producer price index are already made available to the users in A Conjuntura Económica. This publication (available 45 days after the end of the quarter) will be the main channel for disseminating short-term statistics.

Publishing plans:

Sector information (according to ISIC Rev. 2) on the national level.

Historical time series:

Raw data are available from 1997. The INE is aware of the need for time series on relevant user levels.

The construction survey

This is an important sector of the economy – often highly affected by the ups and downs in the business cycle.

Collected data:

Value data on specified construction activities (new and rehabilitation) on residential and non-residential buildings, roads, bridges etc. Monthly.

The survey does not include volume indicators.

Publishing plans:

Information grouped by construction activities and totals for the sector. Data will be released in A Conjuntura Económica.

Historical time series:

Time series are available, but not comparable with the new survey.

Other:

Combining the survey results with other governmental information will be considered, e.g. building permits, floor area information etc.

The hotels and restaurant survey

Collected data:

Production information in value and volume (rooms, number of visitors and number of guest nights) with a split of national and foreign visitors. Monthly.

Actual publishing:

Volume indicators on a national level covering number of visitors and guest nights – national, foreigners and totals – are published in A Conjuntura Económica.

Publishing plans:

The amount of publishing will be increased during this year.

The domestic trade and service sector survey

This survey covers most of the domestic trade (retail, wholesale and sales of cars and petroleum – and the service sectors excluding the financial sector).

Collected data:

Sales values – monthly.

Publishing plans:

Sector information covering retail trade, wholesale trade and a total. Service sector will be published separately. All sector data (according to ISIC Rev. 2) on national level.

Historical time series:

Time series covering the retail trade, wholesale trades are available. The service sector survey is new.

Other:

Combining the survey results with one based on the CPI the data will enable the production of a volume indicator for the retail sector.

The transport survey

This survey covers transports on roads, railway, ships, by air and through pipelines.

Collected data:

Production values and volume indicators (number of passengers, passenger-km, transported goods in ton and km etc.) – monthly.

Publishing plans:

Sector information (according to ISIC Rev. 2) on a national level using breakdowns by types of transport.

Historical time series:

Time series are available for some types of transport.

The harbour and airport survey

This survey covers transports via harbours and airports.

Collected data:

Production values and volume indicators (number of flights, number of embarking/disembarking passengers, national/international flights, Volume (ton) of freight handled, national/international) – monthly.

Actual publishing:

Volume indicators (on national level) for transports by airway (domestic, international and total) and transports through harbour (ton) are published in A Conjuntura Económica.

Historical time series:

Some time series are available.

The external trade statistic

Based on the custom declarations made at the border. Cover all external trade with goods.

Collected data:

Administrative information from the Custom Authorities.

Actual publishing:

Total exports, imports and balance of trade in values. Value indicators (on national level) for important exported and imported goods. Data are published in A Conjuntura Económica.

Historical time series:

Time series are available.

Other:

Due to poor quality in quantity information in the declaration, no unit price information is published.

Appendix C Quality characteristics in Business statistics

The quality of a statistical product might be described by a list of characteristics. The characteristics specified have in common that they express attributes that empirically are known as important to the user.

In addition to the characteristics mentioned we should add the cost of producing the statistical products. This is obvious due to the fact that there is a trade-off between the costs (or the budgets in combination with the priorities of the INE) and the ability to fulfil user requirements. And implicitly there is a bottom line in this list of characteristics, that each user adds up on his own. This concerns the users' credibility to the producer of statistics or the statistical environment responsible for the product.

Building credibility requires clarity and well-functioning routines that are able to produce statistics on a regular basis that meet user needs. Thus – adopting a quality system as a framework for the current production routines is a strategic goal. See also Eurostat (1997) for more about this issue.

Relevance of statistical concepts

In general a survey is relevant if it meets the users' needs.

This focus is on one side on whether the INE has adopted economic concepts for the surveyed variables that are relevant in the user context. If using concepts that are coherent with the National Accounts this will normally be the case for most uses within the national economic area. However, the user's impression of the term relevancy or non-relevancy will often relate to a various number of other issues that are considered important when evaluating the quality of the statistics.

Accuracy in estimates

The accuracy is concerned with the closeness between the estimated value and the (unknown) true population value. Assessing the accuracy of an estimate involves analysing the total error associated with the estimate.

The total errors consist of various types: Sampling errors, non-sampling errors, frame errors, processing errors, non-response errors, measurement errors and model assumption errors. For most surveys the task of making a complete analysis of total errors – in a precise way – is considered as complicated and costly. For most user purposes it is acceptable to provide indirect and partial information on accuracy.

Timeliness and punctuality in dissemination

Most users want up-to-date figures that are published frequently and on time at pre-established dates.

The INEs' role in the information industry is normally being the intermediate producer in the line of users / analysts systemising and presenting information about the economy to the decision-makers. Due to this there exists a close relationship between the INE and the users where advance information about when the products are to be delivered not only seems reasonable, but more or less obvious. Due to the fact that there are several users of the same product, one efficient way to provide advance information is to establish a statistical calendar – updated on a regular basis. Just as important is of course that the products are delivered on time.

How to fulfil the user needs as concerns timeliness is of course a management issue closely related to the resources available and strategic plans made. A typical frequency for short-term statistics is monthly or quarterly. In a situation of scarcity of resources, using mainly a quarterly frequency for sector statistics

allowing for providing a wider coverage of statistics to the user. The use of a quarterly frequency covers most user needs and supports the important Quarterly National Accounts initiative.

Providing the user with advance information on publishing of the statistics has become an important part of the statistical dissemination programme. Such information is especially important for the professional users, which – for the time being – seem to be the typical private user of the INE short-term statistics. For this purpose a statistical calendar should be updated on a regular basis. The calendar must contain pre-established dates for the publication of the statistics.

Accessibility and clarity of the information

Statistical data have most value when they are easily accessible to users, are available in the forms users desire and are adequately documented. Assistance in using and interpreting the statistics should also be available to users.

Firstly, users should be in a position to easily know what kind of statistics that are available. Secondly, physical access to statistics should be convenient. For this purpose A Conjuntura Económica (ACE) in a paper edition might serve many user needs. However, the near future dissemination using the Internet for this type of statistical information seems to be the most convenient and user-adapted solution. Thirdly, the statistics should be accompanied by the necessary information on concepts and methods.

Comparability

Statistics for a given characteristic have the greatest usefulness when they enable reliable comparisons of values taken by the characteristic across space and over time.

As concerns **comparisons across space** the typical user tasks might be to compare e.g. sales between different industrial areas, between provinces, or making comparisons with neighbouring countries.

Comparability should also exist **over time**. Problems in this context might occur due to a number of causes. A common cause might be changes due to modification of the reference concept or changes in the measurement process. The necessary links in the time series should be established (if possible) and the process should be documented and the impact on series from the changes assessed. Even changes in society in general e.g. new legislation, mergers and de-mergers having an impact on continuity in the series should be taken into account.

Coherence

When originating from a single source, statistics are coherent in that elementary concepts can be combined reliably in more complex ways. When originating from different sources, and in particular from statistical surveys of different frequencies, statistics are coherent insofar as they are based on common definitions, classifications and methodological standards. The message that statistics convey to users will then clearly relate to each other, or at least will not contradict each other.

Appendix D The statistical release calendar

The professional analyst spends much time on collecting relevant information for the various types of analyses. Many of the analyses are however of a regular type requiring – more or less – the same type of information every month, quarter – like STS statistics. Such information is especially important for the professional users, which – for the time being – seem to be the typical private user of the INE short-term statistics. For this type of analysis the professional users require advance information on the planned dates for publishing the statistics – to organise the work, make business appointments with his customer etc.

In many countries the distribution of advance information on dates for planned publishing of the various statistics has become an important part of the statistical dissemination programme. This type of advance information has a strategic role in improving the user accessibility to the statistics. An advance date (or week) for releasing each statistics – in a specific month or quarter – known to the user brings clarity for the user's dispositions and planning of his own time. Furthermore – the user credibility improves along with the statistical producer's ability to fulfil the plans made.

It is well-known to most producers that the user requires monthly statistics that are released rapidly – preferably within some few weeks after the period concerned. This is a desire that might be hard or even impossible to fulfil when the resources are scarce. The main practical purpose with the statistical calendar is, however, not to improve speed but – as mentioned above – to bring clarity to the user when planning his own time.

The table below sketches out a statistical calendar – contents and information given – as a frame for further discussions on improving the accessibility to statistics. The structure is fairly simple but contains – in a compact way – most information needed.

Some general aspects

Providing such information requires a system for updating and – of course – discipline and capability to fulfil the targets for publishing as indicated in the calendar.

An important recommendation for such an initiative is though to be extremely realistic. Do not use the introduction of a statistical calendar to change the existing publishing plans. The main target for the initial phase must be – step by step – to enable an integration of the dissemination tasks with the other parts of the production routines – and to succeed in keeping up with the plans.

All statistics published by the INE should be covered by the calendar.

To simplify the access by foreign users the calendar text (name of statistics) should normally contain information in Portuguese as well as in English.

Period

The calendar should be specified for a given period – e.g. 3 months ahead. The calendar should be updated on a monthly basis to always include the new statistics planned for releasing 3 months ahead. This will also make revisions possible when needed.

The name of statistics

The official name of the statistics should be used. If official names for the statistics in English do not exist, it is recommended to create such names. This cell should also include the reference month to be released.

Last and new release

The last and the new releases are specified separately primarily as a reference for the user. If data are available through the Internet, this might simplify the access.

Where/how releasing

This cell should contain information on the main publication series (if any) or channel that is used for releasing the statistics concerned. Often one will find – as indicated in the table below – terms like press release, ACE (A Conjuntura Económica).

Internet

The column Internet is used for indicating if the information is available on the Internet.

Contact person

This cell should contain the name of the contact person, the telephone number and the e-mail address when available. The main purpose is of course to further improve the accessibility of the statistics to the users – and thus the total qualitative value of the product. Such contact will normally also provide an important feedback from various users to the production line – a communication strongly needed.

The last line below the table

The last line (which in fact should be made even more visible to the user by finding another place in the calendar) is included to help users with general requests or requests for ordering statistical publications directly to the Information Centre.

The distribution of the statistical calendar

More important is, however, the dissemination of the statistical calendar to the users. Most efficient is to make the calendar available on the Internet. In a mainly paper-based regime this type of information could be made available on a regular basis as an attachment (or part) of A Conjuntura Económica and in other documents released by the INE. In most European countries the calendar is disseminated to media (papers, radio etc.), which distribute the information to their readers or listeners. All types of dissemination of the calendar should be used.

STATISTICAL CALENDAR. PERIOD: 1.11.2000 - 1.2.2001					
Name of statistics	Last release	New release	Where/how releasing	Internet	Contact person
Consumer Price Index, October 2000	10.10.2000	11.11.2000	Press release	Yes	Name of contact, telephone no., e-mail address
Consumer Price Index, November 2000	11.11.2000	08.12.2000	Press release	Yes	Name of contact, telephone no., e-mail address
Consumer Price Index, December 2000	08.112.2000	10.01.2001	Press release	Yes	Name of contact, telephone no., e-mail address
The Production Index, Manufacturing. III Trim 2000	12.8.2000	11.11.2000	Press release	No	Name of contact, telephone no., e-mail address
Annual statistical yearbook. Demographics - 1998	15.09.1999	15.12.2000	Name of series	No	Name of contact, telephone no., e-mail address
Annual statistical yearbook. Services - 1999	24.12.1999	04.12.2000	Name of series	No	Name of contact, telephone no., e-mail address
A Conjuntura Económica, IV Trim 2000	15.110.2000	15.21.2001	Name of series	Yes	Name of contact, telephone no., e-mail address

For more general information about or how to obtain publications – contact our Information Centre: telephone no., e-mail address.

Appendix E A documentation of the Mozambican CPI (in Portuguese)

This appendix contains a documentation of the Mozambique Consumer Price Index based on a model adopted in Norway for dissemination of information about the statistics published. The appendix is written by Dr. Firmino Guiliche, Head of Department for Prices and Short-term Statistics, DIGCN.

ÍNDICE DE PREÇOS NO CONSUMIDOR BREVES REFERÊNCIAS METODOLÓGICAS

Índice de preços no consumidor

Esta documentação tem em vista colocar à disposição do usuário informação essencial sobre aspectos básicos de compilação do Índice de Preços no Consumidor. Outra informação complementar sobre esta matéria pode ser solicitada no INE na pessoa responsável indicada mais adiante.

Este documento contém informação administrativa; objectivos e finalidade; produção estatística; sobre conceitos básicos, as variáveis e as classificações; tipos de erro e de incerteza; comparabilidade e correlação e; por último, as formas de disseminação.

1. Informação administrativa

1.1. Nome

O Índice de Preços no Consumidor

1.2. Periodicidade

São publicados mensalmente os índices de Maputo, Beira, Nampula e o agregado das três cidades.

1.3. Cobertura territorial

O Índice de Preços no Consumidor compreende a área urbana do País, sendo a recolha e processamento de dados efectuados nas cidades de Maputo, Beira e Nampula.

1.4. Área responsável

INE - DCNIG - Departamento de Preços e Conjuntura, contactar com Firmino Alberto Guiliche – *Chefe do Departamento*

1.5. Legislação

Lei nº 7/96 de 5 de Julho

1.6. Relações internacionais

Em curso acções preparativas para a compilação de um índice harmonizado a nível dos países membros da SADC. Troca de documentação sobre o IPC tanto com os países da SADC como com os PALOP's. Observância das recomendações internacionais sobre a produção de índices, nomeadamente das Nações Unidas e OIT.

2. Objectivo e finalidade

2.1. História e propósitos

O objectivo do IPC é avaliar a evolução dos preços de um conjunto de bens e serviços, de qualidade constante, representativo da estrutura de consumo de uma determinada população, num determinado espaço geográfico

O IPC é publicado mensalmente pelo INE, com dados disponíveis desde Dezembro de 1994 para a cidade de Maputo e Dezembro de 1995, para as cidades da Beira e Nampula. A última revisão do IPC, foi efectuada em Dezembro de 1998 com a actualização do cabaz de bens e serviços e dos ponderadores.

2.2. Utilizadores e aplicações

O índice de preços no consumidor é de utilização ampla. Os bancos utilizam o IPC para monitorar a inflação em ligação com o comércio externo, taxas de câmbio, oferta de dinheiro e taxas de juros; para as contas nacionais ele é usado como deflator; o sector privado utiliza o índice para aumentos de custo ou ajustamento de preços; serve de instrumento de avaliação das políticas de salários e de preços.

3. Produção

3.1. Universo

São considerados os estabelecimentos comerciais e mercados onde a população urbana residente efectua as suas compras. Inclui-se também as entidades de prestação de serviços públicos (saúde, educação, etc.)

3.2. Fonte de dados

Inquérito mensal aos estabelecimentos comerciais e mercados das cidades de Maputo, Beira e Nampula.

3.3. Amostra

A selecção de estabelecimentos e mercados a inquirir é feita na base de critérios bem definidos (postos de venda fixos, volume de vendas apreciável, dispersão geográfica).

Uma componente importante na selecção dos estabelecimentos é a selecção do cabaz e determinação dos ponderadores a partir dos resultados do IAF. Os produtos seleccionados do cabaz são posteriormente identificados nos estabelecimentos a incluir na amostra.

A amostra de estabelecimentos é permanentemente actualizada sempre que as circunstâncias assim o exigirem através de um processo de substituição de estabelecimentos.

3.4. Recolha de dados

A recolha é feita por entrevistadores das delegações do INE por entrevista directa aos estabelecimentos e mercados. Para os produtos frescos, que pelas suas características apresentam variações significativas ao longo do mês a recolha é semanal e para os restantes é mensal.

3.5. Tempo gasto para a recolha.

Em média são 15 minutos de entrevista para os estabelecimentos comerciais e 20 a 30 minutos para os mercados.

3.6. Revisão e controle de qualidade

Os dados dos inquéritos são verificados antes da digitação e confirmados quando necessário nível da delegação do INE respectiva. Após a digitação é feita outra verificação local dos dados de modo a corrigir os erros que possam resultar duma má digitação. Uma verificação definitiva e também efectuada a nível central e em articulação directa com as delegações são efectuadas as últimas correcções.

3.7. Análise de dados

Para a agregação dos índices, adoptou-se a fórmula de Laspeyres, que é um índice ponderado, com ponderações fixas obtidas a partir de um painel de consumo estabelecido para o período de referência. Realiza-se a nível central a um nível mais agregativo e tem a ver com estimações, contribuições e consistência dos dados ao longo do tempo. Habitualmente usam-se para o efeito, gráficos, variações e dados recolhidos por outras instituições

4. Conceitos, variáveis e classificações

4.1. Definição dos conceitos principais

4.2. Definição das variáveis

A variável principal é o preço de transação pago pelos consumidores, que pode ser diferente do recomendado, do oficial ou do estabelecido na lista de preços. São medidos os preços com os quais a maior parte dos consumidores se confrontam, e das quantidades mais típicas.

4.3. Nomenclaturas e Classificações padrão

Adaptação de nomenclaturas e classificações internacionalmente reconhecidas

5. Fontes de erro e de incerteza

5.1. Erros de recolha e processamento

Preços falsos, preços de produtos vendidos em unidades de medida não convencionais, má digitação de dados

5.2. Erros de amostragem

Mudança constante de actividade dos estabelecimentos em que são recolhidos os preços, mudança constante de marcas e qualidade de produtos típicos. Os problemas são minimizados com a actualização permanente da amostra de estabelecimentos e da lista de produtos do cabaz.

Não respostas

Recusa do entrevistado em fornecer o preço, preços de produtos com comportamento sazonal, rotura de stocks

5.3. Erros não amostrais

Erros do registo

Os erros de registo estão minimizados durante o controle e a revisão de qualidade antes da digitação e depois do processamento

6. Comparabilidade e correlação

6.1. Comparabilidade temporal e espacial

A série do IPC data tem como base Dezembro de 1994 para a cidade de Maputo e Dezembro de 1995 para as cidades da Beira e Nampula. Outras séries existiram antes da independência e nos anos 80 mas que não são comparáveis as actuais.

7. Disponibilidade

7.1. Endereço

Os dados sobre IPC estão disponíveis no nosso endereço na Internet: www.ine.gov.mz, e a publicação mensal pode ser adquirida na Av. Ahmed Sekou Touré nº21 – 5º Andar, Flat 55.

7.2. Língua

As tabelas disponíveis da Publicação Mensal do IPC estão em português podendo ser encontradas também no Anuário Estatístico e no Moçambique em números em inglês.

7.3. Publicações

As Publicações Mensais do IPC das cidades de Maputo, Beira, Nampula e o agregado estão disponíveis até ao dia 20 de cada mês.

7.4. Informação de base

O INE tem disponível outra informação de base para o cálculo do IPC, nas delegações provinciais e na sede.

7.5. Outra documentação

Outra documentação adicional encontra-se disponível no INE e delegações:

- Metodologia de calculo
- Relatorios das consultorias 1,2 e 3 do IPC
- Manual da OIT sobre IPC

Appendix F Time schedule for the production line – an example

The table below gives an example on a draft production plan (for main activities) and how this could be specified in writing. The plan sketched out is as such an outcome in writing from a planning process aiming at producing short-term statistics within a 35-40 days delay after the end of month of observation.

Time schedule for the production process of the Manufacturing survey

Ano: 2000

Supervisor and responsible for the statistic: Name

Activities	Persons responsible	1999			
		Dec.	Jan.	Feb.	Month: -----
Maputo CU activities	Supervisor				
Checking of existing questionnaires	Name	18.	21.	19.	
Printing of questionnaires	Name	20.	23.	23.	
Packing and sending questionnaires to the delegations	Name	28.	31.	29.	
Data collection in the delegations	Delegation head				
Questionnaire to the respondents	Name	04.01	01.02	01.03	
Return of questionnaire - respondent, deadline	Name	17.	15.	15.	
Non-response - follow up of <i>critical units</i>	Name				
Start - data entering	Name	21.	22.	22.	
Finish - data entering	Name	22.	23.	23.	
Units not responding - re-contact, last check of data	Name	24.	25.	25.	
Return of key entered data (e-mail) to Maputo CU	Name	26.	27.	27.	
Maputo CU activities	Supervisor				
Data from ministries, EDM, CFM etc.	Name	26.	27.	27.	
Data processing, Maputo CU	Name	02.02	06.03	03.04	
Control of macro results, tables, graphs etc	Supervisor and group	2.	6.	3.	
Tables, graphs for analysis - verification	Production manager	3.	7.	4.	
Final verification	Department head	4.	8.	5.	
Publishing process	Supervisor				
Make final tables, graphs - text commenting recent changes	Name	5.	9.	6.	
Distribution of tables etc. to all internal receivers; internet-distributor; editor of ACE, others	Name	5.	9.	6.	
Date of releasing the statistic	Name	6.	10.	7.	
Person available for user requests	Name				

In other words – drafting a plan is as such not the basic issue when running production lines according to plans based on time schedules. The basic issue is to establish a consistent plan for a given time interval – start of the monthly production process and end of the process (e.g. when results are published). An important part of the planning phase is to assure that all staffs either in the provincial delegations or Maputo CU are familiar with time schedule and the activities involved.

All dates are of course very tentative, but seem fairly reasonable for the normal time needed for producing a monthly short-term statistics. Behind each activity there will be series of minor and larger sub-activities, which require assistance from other persons within the team. All staffs must be familiar with the tasks and aware of the priority given. The plan must of course be developed and encompassed into the INE situation.

Appendix G Critical units in data capture and editing – an example

The term *critical unit* is used for sampled units (establishments) that in advance are known to have a large impact on the estimated results of a level for publishing. The critical units (within a sector aggregate) will have a special focus during data collection and data editing including control of data (micro and macro). A special focus on the critical unit is needed for the management of the sampled units expected to have high influence on the aggregate concerned.

Most critical units will normally be identified in advance based on the relative size of the unit (measured in e.g. sales value, number of employees) within the sector concerned. The task of identifying the critical units is, however, closely related to how the estimation process operates. If the estimations are done on a sector level the critical units should be identified on a sector level. If however, the basic estimations are done within strata of the sector – and afterwards aggregated to a sector level, the critical units should be identified on a sector by stratum level. For some business sectors this is an important distinction – especially when units have been sampled (drawn) in strata having very small average sampling probabilities. In such cases which are likely to occur in sectors having many small units and some large units (e.g. trade sectors), the smaller units will in estimation have a larger impact on the sector aggregate than the larger units.

On some occasions one experiences – not in advance – that some units have large monthly variations in the characteristic observed that actually have an influence on the estimates. If such units repetitively create problems during data editing and estimation, the units concerned should be added to the list of critical units.

If the data collection processes in each province also are expected to produce a data basis for the estimation of provincial statistics, it might be relevant to include the provincial dimension when identifying the critical units.

During data collection the critical unit is focused on simply to avoid a unit non-response. If a unit non-response from the critical unit is a reality, the delegations must provide additional information putting the non-response in a practical context. In general – what is important to know is whether there has been a stop (full or partial – and for how long) or not for the period concerned. This might be information like e.g. a unit has a full or partial stop due to technical problems, vacations etc. This type of information is important when the Maputo staff later on in the production process will consider imputations (part of data editing) for the unit concerned.

The list of critical units in a province is of importance for the delegation's data capture – primarily as a priority list. I.e. which units to concentrate on when having many or too many unit non-responses. The critical unit must also be subject to special treatment during data editing – as concerns the need for re-contact when having a unit non-response, or how to impute when values are missing etc. A special focus on the critical units furthermore implies that more time than average should be spent on assuring that the form is received.

The critical units will in the Maputo CU normally be treated manually by a subject matter specialist.

As the discussion above indicates – the critical unit will be defined using various approaches. The example below is drafted only looking at the sector as such. Some criteria for identifying critical units – in advance – might be:

- ☞ **A critical unit must be identified within a relevant publishing context (frame), e.g. the sector level.**

If the purpose of the survey is to produce reliable statistics on the national level – estimating on a sector level – the sector should be the frame. If the use of the statistics implies that good

quality is required on stratum or provincial sector level as well, the list of critical units must be produced in a multi-dimension setting (e.g. province by sector by stratum).

- ☞ **Every unit having a share of 10 or more per cent of the total value for the sampled units within the relevant sector level should be considered as critical.**

This criterion is based on each unit's share of the total value for the sampled units measured in sales, employment etc. It is recommended to use the number of employees in a unit when estimating the shares. The use of employment should work for all variables covered by the STS questionnaires. The correlation between size of employment and the total value of sales, production values, sum of salaries and wages should normally be high.

Whether the share should be fixed at a 10 per cent level or another level is partly a matter of convenience. This issue will be discussed below – in the light of the next criteria suggested.

The next criterion works on the cumulative distribution of the sampled units. For this purpose the shares for every sampled unit should be estimated – using e.g. employment data. For each of the sectors the units are sorted (in a top-down order) according to the shares. For the group of ranked units the cumulative distribution should be estimated.

- ☞ **All units being ranked below the 75 per cent threshold of the cumulative distribution for the sector should be included.**

This criterion allows for the list of critical units for a sector comprising both medium and large units when the sector has a mixed structure including large and medium sized units.

See next page for a table presenting an excerpt of three sectors from the sampled units in manufacturing for 2000. The list of sampled units given in the table does not include all sampled units for the sector.

Share of sector PV	- the unit share of total production value for the sector.
Cum. share of sector PV	- the cumulative share for the current unit and all units having larger shares within the sector.
Critical units identified	- units fulfilling either one or both of the criteria specified above.

As mention previously in this appendix other units could have been added to the list if taking the provincial dimension into consideration. Units having large monthly variations – not known in advance – should also be considered added to the list when relevant.

Critical units in data capture and editing. Manufacturing survey
An excerpt of units from the 2000 sample

Unit code	ISIC3	Production value	Ranked in all sampl	Share of sector PV	Cum. share of sector PV	Critical units identified
148400	311	96153370	1	0,43	0,43	Critical
155300	311	85750074	3	0,38	0,81	Critical
411700	311	15183405	17	0,07	0,87	
495300	311	5057239	29	0,02	0,90	
157400	311	4111696	32	0,02	0,92	
575900	311	3829271	34	0,02	0,93	
2400	311	3332680	38	0,01	0,95	
1294000	311	2898407	41	0,01	0,96	
895302	311	1853102	51	0,01	0,97	
584800	311	1475036	55	0,01	0,97	
65000	311	1246875	61	0,01	0,98	
6200	311	996000	63	0,00	0,98	
344200	311	995290	64	0,00	0,99	
193800	311	613566	78	0,00	0,99	
253800	311	598064	79	0,00	0,99	
1318700	311	551628	84	0,00	1,00	
110500	311	351128	99	0,00	1,00	
108500	311	337699	101	0,00	1,00	
332900	311	52210	123	0,00	1,00	
98500	311	16800	125	0,00	1,00	
1464700	312	10257000	23	0,98	0,98	Critical
1017800	312	167130	108	0,02	0,99	
117000	312	76184	119	0,01	1,00	
1196007	313	48833998	4	0,32	0,32	Critical
119600	313	45727971	5	0,30	0,63	Critical
1196000	313	26406158	8	0,18	0,80	Critical

An critical unit is identified when having a single share of sector production value larger than SS and/or being among the ranked units having a cumulative share of the sector production value lower than CS, see below.

Cumulative share (cs): 0,8
Single share (ss): 0,1

Appendix H The STS samples

Two tables are presented both giving sample and population information for each of the sectors covered and the 4 strata used. For each stratum a sample probability is estimated to indicate the relative size of the sample. The first table gives data measured in number of units, while the second table provides data measured in number of persons. The sector is defined using 2- or 3-digit ISIC level (rec. 3) groups.

Table 1. Population, samples and sampling probabilities for the STS surveys, by strata.
Grouped by sector (ISIC Rev.3). Number of establishments.

Sector code	Size	Stratum ¹				All strata
		2	3	4	5	
10	Sample	2	1	4	18	25
	Population	18	16	4	18	56
	Sampling probability	0,11	0,06	1,00	1,00	0,45
15	Sample	19	13	15	72	119
	Population	619	183	41	72	915
	Sampling probability	0,03	0,07	0,37	1,00	0,13
18	Sample	5	1	5	39	50
	Population	219	24	13	39	295
	Sampling probability	0,02	0,04	0,38	1,00	0,17
23	Sample	1	2	5	25	33
	Population	20	31	13	25	89
	Sampling probability	0,05	0,06	0,38	1,00	0,37
26	Sample	1	1	2	23	27
	Population	15	19	2	23	59
	Sampling probability	0,07	0,05	1,00	1,00	0,46
28	Sample	3	4	14	43	64
	Population	126	60	24	43	253
	Sampling probability	0,02	0,07	0,58	1,00	0,25
36	Sample	7	11	23	71	112
	Population	226	140	56	71	493
	Sampling probability	0,03	0,08	0,41	1,00	0,23
45	Sample	3	9	29	109	150
	Population	95	115	29	109	348
	Sampling probability	0,03	0,08	1,00	1,00	0,43
50	Sample	10	50	29	34	123
	Population	333	153	29	34	549
	Sampling probability	0,03	0,33	1,00	1,00	0,22
51	Sample	14	65	48	49	176
	Population	454	214	48	49	765
	Sampling probability	0,03	0,30	1,00	1,00	0,23
52	Sample	90	58	45	17	210
	Population	5211	358	45	17	5631
	Sampling probability	0,02	0,16	1,00	1,00	0,04
551	Sample	15	92	22	29	158
	Population	72	92	22	29	215
	Sampling probability	0,21	1,00	1,00	1,00	0,73
552	Sample	60	90	28	8	186
	Population	1243	220	28	8	1499
	Sampling probability	0,05	0,41	1,00	1,00	0,12

¹ Stratum grouped by number of employees: Above 50 (5), 36 - 49 (4), 20 - 29 (3) and 1 - 19 (2)

**Table 1 (cont.). Population, samples and sampling probabilities for the STS surveys, by strata.
Grouped by sector (ISIC Rev.3). Number of establishments.**

Sector code	Size	Stratum ¹				All strata
		2	3	4	5	
601	Sample	2	3		15	20
	Population	2	3		15	20
	Sampling probability	1,00	1,00		1,00	1,00
602	Sample	14	15	11	21	61
	Population	88	52	11	21	172
	Sampling probability	0,16	0,29	1,00	1,00	0,35
603	Sample		3		2	5
	Population		3		2	5
	Sampling probability		1,00		1,00	1,00
611	Sample	5	11		5	21
	Population	15	11		5	31
	Sampling probability	0,33	1,00		1,00	0,68
62	Sample	12	3	4	3	22
	Population	13	10	4	3	36
	Sampling probability	0,92	0,30	1,00	1,00	0,61
63	Sample	11	12	10	22	55
	Population	62	36	10	22	136
	Sampling probability	0,18	0,33	1,00	1,00	0,40
64	Sample	4	7	5	15	31
	Population	140	21	5	15	181
	Sampling probability	0,03	0,33	1,00	1,00	0,17
65	Sample	5	20	10	19	54
	Population	60	63	10	19	152
	Sampling probability	0,08	0,32	1,00	1,00	0,36
70	Sample	2	4	4	6	16
	Population	8	4	4	6	22
	Sampling probability	0,25	1,00	1,00	1,00	0,73
72	Sample	2	3			5
	Population	4	4			8
	Sampling probability	0,50	0,75			0,63
79	Sample	16	10	28	32	86
	Population	536	150	28	32	740
	Sampling probability	0,03	0,07	1,00	1,00	0,12
80	Sample	1	15	12	11	39
	Population	55	45	12	11	123
	Sampling probability	0,02	0,33	1,00	1,00	0,32
All industries	Sample	364	503	353	688	1848
	Population	9628	2027	438	688	12781
	Sampling probability	0,04	0,25	0,81	1,00	0,14

¹ Stratum grouped by number of employees: Above 50 (5), 36 - 49 (4), 20 - 29 (3) and 1 - 19 (2)

**Table 2. Population, samples and sampling probabilities for the STS surveys, by strata.
Grouped by sector (ISIC Rev.3). Number of employees**

Sector code	Size	Stratum ¹				All strata
		2	3	4	5	
10	Sample	10	17	183	2669	2879
	Population	86	272	183	2669	3210
	Sampling probability	0,12	0,06	1,00	1,00	0,90
15	Sample	72	216	555	26791	27634
	Population	2348	3034	1518	26791	33691
	Sampling probability	0,03	0,07	0,37	1,00	0,82
18	Sample	18	16	191	10396	10621
	Population	785	375	496	10396	12052
	Sampling probability	0,02	0,04	0,39	1,00	0,88
23	Sample	5	33	183	3599	3820
	Population	102	506	475	3599	4682
	Sampling probability	0,05	0,07	0,39	1,00	0,82
26	Sample	5	16	74	5635	5730
	Population	76	307	74	5635	6092
	Sampling probability	0,07	0,05	1,00	1,00	0,94
28	Sample	14	76	555	4686	5331
	Population	583	1137	951	4686	7357
	Sampling probability	0,02	0,07	0,58	1,00	0,72
36	Sample	33	183	877	8706	9799
	Population	1051	2332	2135	8706	14224
	Sampling probability	0,03	0,08	0,41	1,00	0,69
45	Sample	16	167	1098	21837	23118
	Population	491	2137	1098	21837	25563
	Sampling probability	0,03	0,08	1,00	1,00	0,90
50	Sample	47	826	1099	4207	6179
	Population	1572	2528	1099	4207	9406
	Sampling probability	0,03	0,33	1,00	1,00	0,66
51	Sample	70	1041	1817	6672	9600
	Population	2270	3428	1817	6672	14187
	Sampling probability	0,03	0,30	1,00	1,00	0,68
52	Sample	296	847	1656	1776	4575
	Population	17147	5228	1656	1776	25807
	Sampling probability	0,02	0,16	1,00	1,00	0,18
551	Sample	89	1528	827	2901	5345
	Population	426	1528	827	2901	5682
	Sampling probability	0,21	1,00	1,00	1,00	0,94
552	Sample	222	1391	1027	535	3175
	Population	4595	3399	1027	535	9556
	Sampling probability	0,05	0,41	1,00	1,00	0,33

¹ Stratum grouped by number of employees: Above 50 (5), 36 - 49 (4), 20 - 29 (3) and 1 - 19 (2)

**Table 2 (cont.). Population, samples and sampling probabilities for the STS surveys, by strata.
Grouped by sector (ISIC Rev.3). Number of employees**

Sector code	Size	Stratum ¹			All strata	
		2	3	4		5
601	Sample	11	85		32779	32875
	Population	11	85		32779	32875
	Sampling probability	1,00	1,00		1,00	1,00
602	Sample	63	252	427	3452	4194
	Population	398	872	427	3452	5149
	Sampling probability	0,16	0,29	1,00	1,00	0,81
603	Sample		45		479	524
	Population		45		479	524
	Sampling probability		1,00		1,00	1,00
611	Sample	19	153		1035	1207
	Population	57	153		1035	1245
	Sampling probability	0,33	1,00		1,00	0,97
62	Sample	46	44	132	1844	2066
	Population	50	146	132	1844	2172
	Sampling probability	0,92	0,30	1,00	1,00	0,95
63	Sample	53	203	391	3383	4030
	Population	300	609	391	3383	4683
	Sampling probability	0,18	0,33	1,00	1,00	0,86
64	Sample	12	119	195	3569	3895
	Population	426	357	195	3569	4547
	Sampling probability	0,03	0,33	1,00	1,00	0,86
65	Sample	32	336	392	5653	6413
	Population	380	1059	392	5653	7484
	Sampling probability	0,08	0,32	1,00	1,00	0,86
70	Sample	8	89	129	3759	3985
	Population	31	89	129	3759	4008
	Sampling probability	0,26	1,00	1,00	1,00	0,99
72	Sample	13	80			93
	Population	26	106			132
	Sampling probability	0,50	0,75			0,70
79	Sample	67	152	994	7185	8398
	Population	2209	2273	994	7185	12661
	Sampling probability	0,03	0,07	1,00	1,00	0,66
80	Sample	6	257	482	804	1549
	Population	313	772	482	804	2371
	Sampling probability	0,02	0,33	1,00	1,00	0,65
All industries	Sample	1227	8172	13284	164352	187035
	Population	35733	32777	16498	164352	249360
	Sampling probability	0,03	0,25	0,81	1,00	0,75

¹ Stratum grouped by number of employees: Above 50 (5), 36 - 49 (4), 20 - 29 (3) and 1 - 19 (2)

Appendix I List of contacts

INE:

Valeriano Levene	Vice President, Division for Economic Statistics
Firmino Alberto Guiliche	Head of Department for Prices and Short-term Statistics, DIGCN
José Luis	IPC, Department for Prices and Short-term Statistics, DIGCN
Jeremias Guambe	IPC, Department for Prices and Short-term Statistics, DIGCN
Jeremias Cherinza	A Conjuntura Económica, Department for Prices and Short-term Statistics, DIGCN
António Lazo	Advisor, National Accounts, DIGCN
Margarida Ferreira	LDB, DIGCN
Camilo Amade	Head of Department for Services, Informal sector and Business Register, DESE
Cirillo Tembe	Business Register, DESE
António Júnior	Head of Department for Goods and Environment, DESE
Francisco Nuvunga	Manufacturing Industry, DESE
Maria Teresa Tovela	Manufacturing Industry, DESE
Marcelino Silva	Head of Department for Dissemination and Distribution, DICRE
Leif Korbøl	Statistical advisor, Twinning Arrangement, Economic Statistics, INE
Leif Norman	Team leader, Twinning Arrangement, INE
Bo Yttergren	Statistical Advisor, Twinning Arrangement, IT, INE

DPINE Maputo Cidade

Remingo Bernabe Nkondya	Head of Department for Economic Statistics, Goods and Services
Jorge Daniel Chemane	Supervisor, CPI sector
Rosalina C. Machatine	Head of Department for Administration and Human resources

DPINE Maputo Província

Ibraimo Ali
Virgínia Bohane
António Melice

Ministerio de Plano e Financas

Pedro Couto	Director, Gabinete de Estudos, MPF
Bruce R. Bolnick	Advisor, Gabinete de Estudos, MPF
Per-Åke Andersson	Advisor, Gabinete de Estudos, MPF

Ministério de Indústria

Elsa

World Bank

Sherry Archondo	International Finance
-----------------	-----------------------

The Swedish Embassy

Maude Svensson	Economic Advisor, Country Economist
----------------	-------------------------------------

Appendix J List of Literature

Solheim and Richardsson (2000), Industry Statistics in Mozambique, Documents 2000/8, Statistics Norway

Johansson and Åstrøm (2000), Printing and Dissemination, Report from the Twinning Project (draft)

Wold (2000), Planned Co-operation with Instituto de Estatística (INE), Mozambique, Documents 2000/10, Statistics Norway

Eurostat (1997), Quality in Business Statistics

Recent publications in the series Documents

- 99/4 K. Rypdal and B. Tornsjø: Construction of Environmental Pressure Information System (EPIS) for the Norwegian Offshore Oil and Gas Production
- 99/5 M. Søberg: Experimental Economics and the US Tradable SO₂ Permit Scheme: A Discussion of Parallelism
- 99/6 J. Epland: Longitudinal non-response: Evidence from the Norwegian Income Panel
- 99/7 W. Yixuan and W. Taoyuan: The Energy Account in China: A Technical Documentation
- 99/8 T.L. Andersen and R. Johannessen: The Consumer Price Index of Mozambique: A short term mission 29 November – 19 December 1998
- 99/9 L-C. Zhang: SMAREST: A Survey of Small Area ESTimation
- 99/10 L-C. Zhang: Some Norwegian Experience with Small Area Estimation
- 99/11 H. Snorrason, O. Ljones and B.K. Wold: Mid-Term Review: Twinning Arrangement 1997-2000, Palestinian Central Bureau of Statistics and Statistics Norway, April 1999
- 99/12 K.-G. Lindquist: The Importance of Disaggregation in Economic Modelling
- 99/13 Y. Li: An Analysis of the Demand for Selected Durables in China
- 99/14 T.I. Tysse and K. Vaage: Unemployment of Older Norwegian Workers: A Competing Risk Analysis
- 1999/15 L. Solheim and D. Roll-Hansen: Photocopying in Higher Education
- 1999/16 F. Brunvoll, E.H. Davila, V. Palm, S. Ribacke, K. Rypdal and L. Tangden: Inventory of Climate Change Indicators for the Nordic Countries.
- 1999/17 P. Schøning, M.V. Dysterud and E. Engeliën: Computerised delimitation of urban settlements: A method based on the use of administrative registers and digital maps.
- 1999/18 L.-C. Zhang and J. Sexton: ABC of Markov chain Monte Carlo
- 1999/19 K. Flugsrud, W. Irving and K. Rypdal: Methodological Choice in Inventory Preparation. Suggestions for Good Practice Guidance
- 1999/20 K. Skrede: Gender Equality in the Labour Market - still a Distant Goal?
- 1999/21 E. Engeliën and P. Schøning: Land Use Statistics for Urban Settlements: Methods based on the use of administrative registers and digital maps
- 1999/22 R. Kjeldstad: Lone Parents and the "Work Line": Changing Welfare Schemes and Changing Labour Market
- 2000/1 J.K. Dagsvik: Probabilistic Models for Qualitative Choice Behavior: An Introduction
- 2000/2 A. Senhaji: "An Evaluation of some Technology Programs executed by the Norwegian Government in the 80's and the 90's
- 2000/3 K. Rypdal and B. Tornsjø: Environmental Pressure Information System (EPIS) for the Pulp and Paper Industry in Norway
- 2000/4 K. Rypdal and B. Tornsjø: Chemicals in Environmental Pressure Information System (EPIS)
- 2000/5 R. Ragnarsøn: The Role of Subcontracting in the Production System
- 2000/6 K.E. Rosendahl: Industrial Benefits and Costs of Greenhouse Gas Abatement Strategies: Applications of E3ME: Modelling external secondary benefits in the E3ME model
- 2000/7 G.A. Ellingsen, K.E. Rosendahl and A. Bruvoll: Industrial Benefits and Costs of Greenhouse Gas Abatement Strategies: Applications of E3ME: Inclusion of 6 greenhouse gases and other pollutants into the E3ME model
- 2000/8 R. Ragnarsøn and L. Solheim: Industry Statistics in Mozambique: Major Findings and Recommendations
- 2000/9 R. Johannessen: The Consumer Price Index of Mozambique: A Short Term Mission 13-31 March 2000
- 2000/10 B.K. Wold: Planned Co-operation with Instituto Nacional de Estatística (INE), Mozambique: Report from Short Term Identification Mission 27th March to 3rd April, 2000 Requested by NORAD/Oslo
- 2000/11 P. Boug: Modelling Energy Demand in Germany: A Cointegration Approach
- 2000/12 E. Engeliën and P. Schøning: Land use statistics for urban settlements
- 2000/13 M. Rønsen: Impacts on Women's Work and Child Care Choices of Cash-for-Care Programs
- 2000/14 H.C. Bjørnland: VAR Models in Macroeconomic Research

Statistics Norway
P.O.B. 8131 Dep.
N-0033 Oslo

Tel: +47-22 86 45 00
Fax: +47-22 86 49 73

ISSN 0805-9411

