Manual on Surveys of Informal Employment and Informal Sector

Draft Chapter 6:

Mixed surveys using the modular approach

January 2010

1. Introduction

As mentioned in Chapter 3, mixed surveys (household/enterprise) measuring the informal sector constitute a broad family which includes a variety of surveys sharing a common denominator. In this manual two main types of mixed surveys are distinguished: independent mixed surveys (Chapter 7) and modular mixed surveys, which are the subject of the present chapter. The objective of the present chapter is to draw a comprehensive picture of statistical properties of this type of surveys, to provide relevant informal sector and informal employment indicators consistent with international definitions and in relation with other alternatives.

As other parts of this manual, this chapter does not claim to cover exhaustively the modular mixed surveys approach. Nevertheless it aims at explaining its rationale and generic design, while underlying the main strong points and shortcomings of the approach. The chapter is based on selected significant and illustrative examples drawn from national experiences.

The chapter is organized as follows. Section 2 presents the main principles of modular mixed surveys. Section 3, the core of this chapter, focuses on the *1-2-3 survey*, which has become the most widely known survey of this kind. The discussion includes the methodology of the surveys conducted under the UN Development Account Project 'Interregional cooperation on the measurement of informal sector and informal employment' in several countries. Given their specificities and high level of institutionalization, the Mexican and South African experiences are also detailed in Section 3. Section 4 is dedicated to reviewing the potential of the *Living Standards Measurement Studie*s to measure the informal sector. The last section (Section 5) draws some perspectives for future developments.

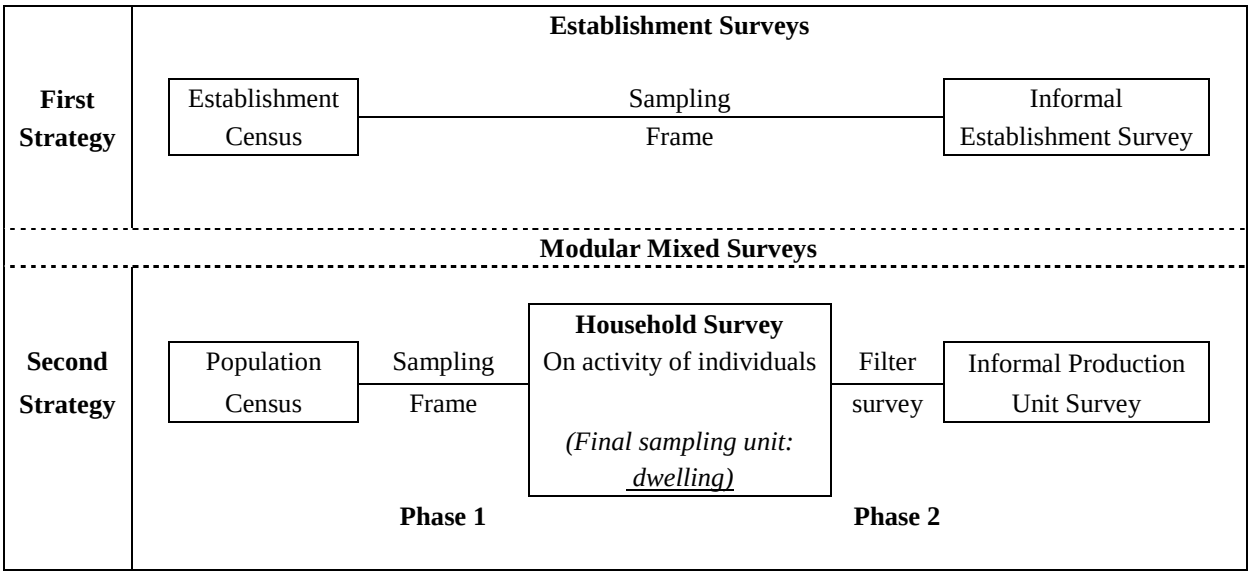
2. Principles of modular mixed informal sector surveys

The general principle of the modular mixed surveys is to use information drawn from a survey of households concerning the activity of individuals (phase 1) to select a sample of owners of production units, to which a specific questionnaire on informal activity is applied (phase 2; Figure 6.1). Each individual belonging to the employed population (for example, any individual who has worked for at least one hour during the reference week, to take the ILO definition), who says that he or she is an employer or an own-account worker in a unit satisfying the criteria used for defining the informal sector (size or non-registration), is asked to complete the questionnaire on the informal production unit he or she is holding.

The two phase design has been developed during the 1980s to try to overcome the main problem of the conventional approach (referred to as "first strategy" in Figure 6.1), i.e. the difficulty in achieving an exhaustive coverage of the informal sector (Roubaud and Séruzier, 1991; Roubaud, 1997). The two major risks with the establishment surveys is, first, to get a comprehensive picture of the reference population (informal production enterprises). Unless an economic census is conducted and articulated with the population census (as for example in India; see Chapter 5), an exhaustive coverage of the universe cannot be achieved, especially for activities performed at home or without fixed location. Second, given the erratic demographic laws (birth and mortality rates) of the informal sector enterprises, updating registers to draw representative samples of informal sector units is usually out of hand. That is why most of the surveys conducted along this line have produced unreliable and non-consistent estimators (underestimating totals; overestimating the weight of informal

production units with premises, and consequently overstating the economic performances of the informal sector, as enterprises with a fixed location usually perform better than those without fixed location.

Figure 6.1: Two sampling strategies for measuring the informal sector



Source: Roubaud and Séruzier, 1991.

The fundamental rationale of informal sector mixed surveys is that informal production units (IPUs)[1] are easier to catch through the jobs of the persons working in the informal sector than through the identification of premises in which the activity is held. The main challenge is then to build the implicit population of individuals, jobs and production units, and their relations. The sequence is, departing from the population, to get to production units through jobs (which play the role of an intermediation variable).

2.1 Sampling issues

The sample design of the modular mixed surveys consists of adding an additional stage of selection to the phase 1 (which is often already a multi-stage sample survey). For example, when the phase 1 is a two-stage LFS, a third stage is grafted to select the sub-sample of heads of IPUs (informal production units). It relies on two assumptions:
- a sufficiently large sample of IPU heads can be drawn from phase 1;
- a bi-univocal relation can be established between the IPU head (identified in phase 1) and his or her production unit (to be surveyed in phase 2).

2.1.1 Building the reference population of IPUs

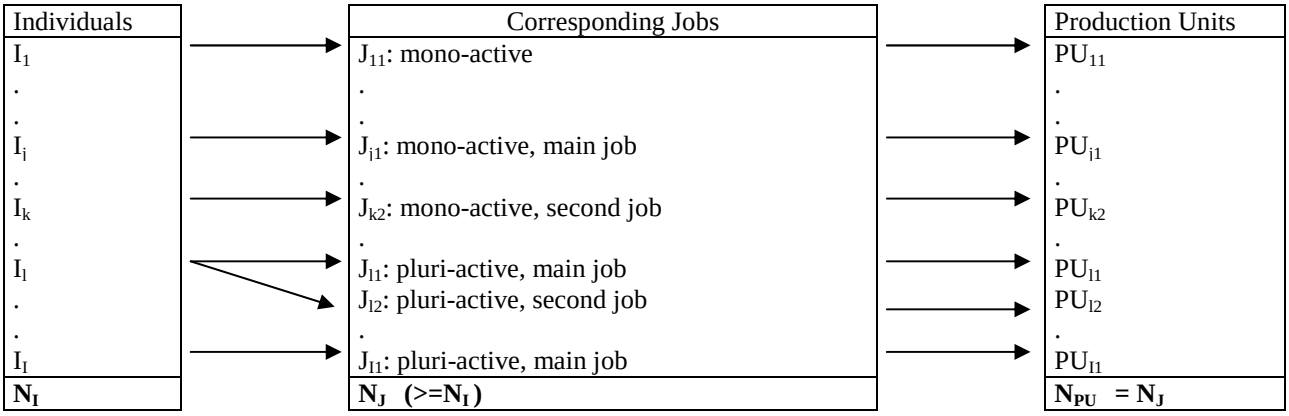
Let us discuss the necessary properties of phases 1 and 2 to ensure the realization of these assumptions. Departing from a representative sample of the households, the first condition to be respected is that information on all jobs held in the informal sector and corresponding to

[1] Given the specificity of "enterprises" in the informal sector (e.g. a tailor repairing shirts at home for neighbourhood clients and working alone only a few hours a week is considered as an "enterprise") this chapter uses the term "production units" rather than "enterprises" or "establishments".

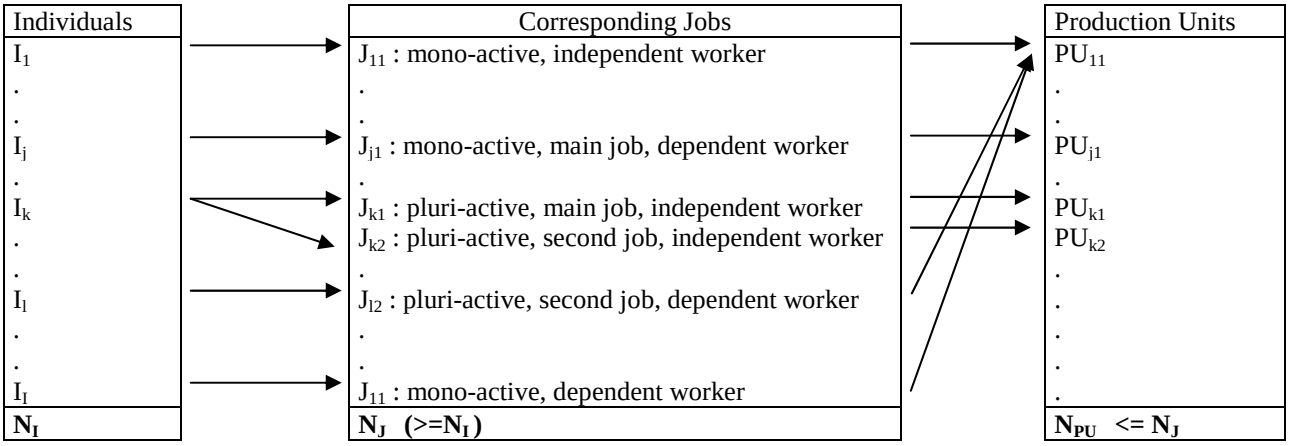
this initial population is made available in phase 1. This means that information about the informal sector of employment should be asked for both the main and secondary jobs.[2] As already mentioned in Chapter 4, appropriate questions for the identification of employment in the informal sector should be addressed to all employed persons irrespective of their status in employment (including employees and contributing family workers), and not only to employers and own-account workers. This makes it possible to obtain an estimate of total employment in the informal sector on the basis of data collected during phase 1, which can be compared to the estimate of total employment in the informal sector obtained from phase 2 on the basis of the information provided by informal employers and own-account workers on the number and characteristics of the persons employed in their IPUs. Experience has shown that the estimate of total employment in the informal sector obtained from phase 1 tends to be higher than the corresponding estimate obtained from phase 2, as there may be reasons for

Figure 6.2: Building the implicit population of individuals, jobs and production units

A.- Selecting only independent workers (employers, own account workers), in the informal sector



B.- Selecting all kind of workers (employers, own account workers, employees, family workers), in the informal sector



Source: adapted from Roubaud and Séruzier, 1991.

[2] In theory, we should consider all the existing jobs during the reference period. Some surveys do. In practice, the number of simultaneous jobs reported for one person during the reference period (usually one week) only exceptionally exceeds two.

informal sector entrepreneurs to underreport the number of persons working for them. Also, because of financial or other constraints it may sometimes not be actually feasible to conduct the phase 2 survey even though it had been planned.

Once the reference population of informal sector jobs built, the problem is to get to the corresponding population of IPUs. Two alternative options can be considered according to the status in employment: select only independent workers (employers, own-account workers); or select all workers (independent and dependent) irrespective of their status in employment. The first option is quite simple to address as there is a "quasi" equivalence between production units and their holders. The second is more complex as different jobs can lead to the same production unit (Figure 6.2).

In the first option, the only violation of assumption 2 occurs in the case of business partners residing in different households. To solve this problem, information is needed on the number of business partners. It can be obtained either in phase 1 (more complicated) or *ex post* in phase 2 (information collected anyway). The initial sampling weight of the business partners' IPUs can be corrected proportionally to (i.e. divided by) the number of business partners residing in different households[3], keeping in mind that IPUs with two or more business partners usually represent only a very small part of the informal sector in practice. Their share may however be larger in specific branches of economic activity requiring relatively high capital investments (e.g. mechanised transport). In general, where the number of IPUs operated in business partnership is significant, more sophisticated methods for the correction of sampling weights should be used (see Chapter 7), which require information on the location of the households of all the business partners. As an alternative, one may identify the main business partner and select only him or her for inclusion in the phase 2 sample.

In the second option, there is a systematic violation of assumption 2 for all IPUs which employ two or more persons residing in different households. The probability of inclusion is proportional to the IPU's size. Nevertheless, if this option is feasible in theory, it is not recommendable in practice for various reasons and, hence, has been rejected by the 15th ICLS (see paragraph 26 (2) of its resolution). First, the formulae of extrapolation coefficients are more complicated than for the first option. Second, employees may have more difficulties in providing reliable information about the informal status of their enterprise (either defined by the size or the legal status). If the sample of IPUs is drawn only from the responses given by owners of IPUs, who know the legal status of their establishment, the problem is solved. Finally, this strategy can raise ethical issues. If the head of an IPU happens to know that his or her production unit has been selected for phase 2 thanks to the declaration of one of his or her employees surveyed in phase 1, he or she can feel resentful and even dismiss the employee. Guaranteeing confidentiality can be problematic. Fortunately, to our knowledge option 2 has never been applied in modular mixed informal sector surveys.

2.1.2 Sample size and stratification

In a typical phase 1 survey, the sampling frame follows a two-stage (enumeration areas, dwellings) stratified (at EA level: generally administrative boundaries) scheme. Despite their good statistical properties for informal sector measurement purposes, two-stage surveys present a number of difficulties which must be taken into account. As, by construction, the

[3] It should be noted that the sampling weights for information, that relates to the personal characteristics of the business partners themselves, should remain unchanged.

phase 1 (properly applied) provides consistent – unbiased – estimates, the sampling problem is to obtain efficient estimators, that is to say the ones which minimize their variance.

The first limitation occurs when the sample size in phase 1 is too small to obtain reliable estimates in phase 2 (according to the desired level of disaggregation of the results – typically by industries or by geographical zones). Such a case can occur when the phase 1 sample has been designed not taking into account the objectives of the phase 2 survey, or when the overall density of informal sector activities is low. The second limitation can occur in the case of marginal and/or geographically concentrated activities. Such is the case in particular of branches of economic activity where there are few informal production units, or of branches concentrated in specific geographical locations. In some countries, craft trades are traditionally grouped in clearly identified places (jeweller streets, tailors, kebab sellers, etc.).

Three strategies can be adopted and combined. The first consists of increasing the sample size for phase 1, if resources permit. A crucial ratio to get *ex ante* is the (approximate) number of IPUs per household. Usually such a ratio is not available (getting it is one of the reason to conduct the survey), but a proxy can be obtained from previous surveys (for instance, the number of independent workers by household, the number of micro-enterprises, etc.).

The second strategy consists of using auxiliary information for additional sample stratification in phase 1. To avoid informal activities not being properly represented, all the information already available should be mobilised and introduced into the sampling frame as a stratification variable. In particular, when selecting sample areas (EAs) it is recommended that account be taken of the density of informal activities and that high-density areas be over-represented in the sample in order to improve the accuracy of the estimators and reduce survey costs. This information may be an outcome of previous surveys or of prior research work.

The third strategy only applies to phase 1 surveys that are conducted at regular intervals, such as continuous labour force surveys. In this case, subsamples of IPU heads can be accumulated over several survey rounds until a sufficiently large number of IPU heads is obtained (cf. paragraph 31(2) of the 15th ICLS resolution). This would normally imply that the periodicity of the phase 2 survey follows that of the phase 1 survey.

Let us now consider the phase 2. The same type of solutions can be adopted to improve the precision of estimators (sample size, stratification). Increasing the sample size is constraint by the total number of IPUs given by the phase 1. At this level, and not taking into account cost issues, the best sampling design for phase 2 is to survey exhaustively all IPUs identified in phase 1. Such a sampling design also presents the advantage of simplifying greatly the calculation of extrapolation weights, variance and confidence intervals (as the probability of selection of IPUs at the last stage is equal to 1). Once the decision has been taken to select only a sub-sample of the IPUs identified through phase 1, it is advisable to stratify the sample using information collected in phase 1. The most common variables for stratification are: industry, status in employment (employer, own-account worker), type of job of the IPU head (main vs. second job), gender of the IPU head, type of premises, income of the head, etc. It is worth highlighting that the stratification procedures are particularly efficient as the characteristics of jobs captured in phase1 are strongly correlated with the characteristics of the IPUs.

Finally, to extrapolate the phase 2 results, non-response (total or partial) should be taken into account, as well as misclassification due to measurement errors in phase 1. A procedure of post-stratification can be implemented to recalculate the final weights accordingly.

2.2 Measurement and logistical issues

Once the sample has been designed, a number of important issues should be addressed to optimize the reliability of the information. This subsection concentrates on four main points: the choice of the filter survey, the time span between phases 1 and 2, the place of interview for the phase 2 survey, and the relevant respondent. The specificities of the survey questionnaires will be discussed in detail in Section 3.

One important measurement issue is the *choice of the filter survey*. Given what has been said previously, particularly the tight links between jobs and IPUs, labour force surveys (LFSs) are the best candidate to serve as phase 1 for a modular mixed informal sector survey. Of course, they have to be adapted in sampling and questionnaire design to the specific objective of measuring the informal sector. Another advantage of the LFSs is to catch two birds with one stone: they are not only the best filter survey to select the phase 2 sample, but can also provide good estimates of employment in the informal sector and in other forms of informal employment (see Chapter 4). However, LFSs are not the only possibility to serve as phase 1. In fact, all types of household surveys providing information on jobs or sources of labour income are potential candidates. For example, LSMS surveys or household income and expenditure surveys (HIES) can provide a sample frame for phase 2 provided the sample sizes are sufficiently large (see Section 5). Another solution is to graft an identifying specific module on any representative household survey (migration survey, victimization survey, etc.). In some countries, where such surveys are planned, grafting such a module can present the economic interest of just paying the marginal cost, and the analytic strongpoint of allowing cross tabulation of the information collected in the module with the one collected in the main survey. Nevertheless, there is a risk of overloading and "contaminating" the original survey, which shall be avoided. Moreover, the sample design requirements for the two surveys may not be compatible.

A second issue is *the management of the time lag between phases 1 and 2*. This time lag should be reduced to the minimum. The longer is the time lag between phase 1 and phase 2, the higher is the risk of attrition, as a higher proportion of IPUs will have disappeared due to high mobility and turnover. In the example shown in column 3 of Figure 6.9 in Section 3.3 of this chapter, even within a time span of only two weeks between the two survey phases the loss rate of IPUs already amounts to around 5 %. The extreme solution to avoid the problem of attrition is to conduct the two phases at the same time (conduct the phase 2 once the phase 1 has been completed and the IPUs identified), or even to integrate the two phases in the same questionnaire (see the LSMS case described in Section 4). Nevertheless this strategy can create problems. First, it is impossible to select a sub-sample of IPUs; the IPUs have to be surveyed exhaustively. Second, the phase 2 survey will take place in the household dwelling although it is preferable to conduct the survey in the IPU's premise (see below). Third, the identification of IPUs for the phase 2 survey is left to the field staff while it should be under the control of the head office. Fourth, both survey phases will have to be conducted by the same team of interviewers while each of them may require a specialized team. The ideal time elapsed between the two survey phases is a trade off between the objective of reducing attrition and the time necessary to select the phase 2 sample (data entry, selection procedures, etc.), to optimize the organization of the enumerators, etc. In any case, the

attrition rate has to be taken into account to obtain consistent estimators. In particular, the phase 2 shall be conducted for all IPUs selected, whether they are still operating or not, provided the IPU heads can still be traced. The reference period should thus be the last operating month which, by definition, is posterior or equal to the month of phase 1. To compensate for attrition due to loss of IPUs, other contact failures, or refusals to participate in phase 2, the phase 1 sample should be increased from the outset by the expected amount of attrition due to these factors.

The *place of interview for the phase 2 survey* is also of high importance. To enhance the reliability of the data collected, it is preferable to conduct the survey in the premises of the IPUs, as in other enterprise/establishment surveys. Three cases should be distinguished. For IPUs located at home or without fixed location, the interview should be conducted at the home of the IPU head. The direct measurement of informal activities at the place of work is inappropriate for itinerant units or activities performed on the public sidewalk. The street is certainly not a suitable place for completing a survey questionnaire. When a fixed location out of home exists, it is most suitable to conduct the survey there. By this means, the interviewers can check the accuracy of the collected information (reducing under-declaration); he or she can also interview directly the other members of the IPU, if the head is not able to provide information about their characteristics (education, training, migration status, etc.). Of course, this strategy raises logistical difficulties (locating the premise, etc.) which need to be addressed.

While LFS data often have to be obtained from proxy respondents, it is a characteristic of the phase 2 survey that *proxy response should not be allowed*. The only person supposed to answer phase 2 is the head of IPU. This means that information collected in phase 2 is in general more reliable than for phase 1, with the possible exception of information on employment in the informal sector. Of course, not accepting proxy respondents has a cost, as the self-respondent must be tracked. However, this has another favourable consequence. Common information collected both in phases 1 and 2 (data on employment status of the IPU head, industry, etc.) can be double-checked and corrected in the LFS, if necessary.

One last point to be mentioned is the need *to design adapted questionnaires* to fit the specific rhythm of activities and the concepts understood by informal sector entrepreneurs. This will be discussed in more detail in Section 3 below.

3. The modular approach and its uses in different contexts

3.1 The 1-2-3 survey

3.1.1 Historical background

Initially designed at the beginning of the 1990s to study the informal sector (Roubaud 1992), the *1-2-3 survey* was gradually extended to measure and monitor also poverty and governance, adapting itself to the increasing importance of these issues, which now constitute the heart of development policy. After an initial partial experiment in Mexico (1987, 1989), the *1-2-3 survey* was applied for the first time in its entirety in Yaoundé (Cameroon) in 1993 (Roubaud 1994b). The methodology was then consolidated in Antananarivo (Madagascar), where the mechanism was introduced in 1995 and in operation until 2006. Initially limited to the capital, it was extended in 2000 after five years of successful operation to the country's

Figure 6.3: 1-2-3 Surveys 1987-2007*

Countries	Date	Sample size			Coverage
		Phase 1 (Household)	Phase 2 (UPI)	Phase 3 (Household)	
Africa					
Benin	2001	3,000	1,000	600	Cotonou
	2003 & 2004	3,600	1,200	No	Urban Areas
	2006 & 2007	18,000	6,200	18,000	National
Burkina Faso	2001	2,500	1,000	1,000	Ouagadougou
Burundi	2006	1,600	600	900	Bujumbura
	2007	3,300	600	No	Urban Areas
Côte d'Ivoire	2002	2,500	1,000	No	Abidjan
Mali	2001	2,400	1,000	1,000	Bamako
Morocco*	1999/2000	45,000	8,900	15,000	National
Niger	2002	2,500	800	600	Niamey
Senegal	2002	2,500	1,000	600	Dakar
Togo	2001	2,500	1,000	600	Lome
Madagascar	1995-2006	3,000	1,000	600	Antananarivo
Cameroon	1993	2,500			Yaounde
	2005	8,500	5,100	No	National
Dem. Rep. of Congo	2004-2005	13,700	6,300	13,700	National
Latin America and Caribbean					
Colombia	2001	48,000	9,000	1,700	Urban Areas
El Salvador	1992				
Haiti	2007/08	8,100	4,400	5,200	National
Mexico	1987			No	Mexico
	1988/89		2,750	No	7 largest Cities
	1992-2002, 2008			No	Urban, 2008 National
Peru	1993		4,300	No	Urban Areas
	2001 & 2002	18,000	8,000	18,000	National
	2003-2007 (continuous)	20,000 (1,600/ month)	12,000	20,000	National
Venezuela	2000	7,600	521	No	National
Asia					
China	2002	3,600	?	3,600	Urban (3 pilot cities)+ rural
Mongolia	2007/08	5,000	2,200	No	National
Vietnam	2007	173,000	3,000	No	National (LFS)

***Notes**: The table only includes countries where at least two phases of the 1-2-3 survey were conducted. In Morocco, the phases 1 and 2 on one side and the phase 3 on the other were disconnected. The first two phases were drawn from the LFS, while the third phase corresponded to the 2000/2001 HIES.

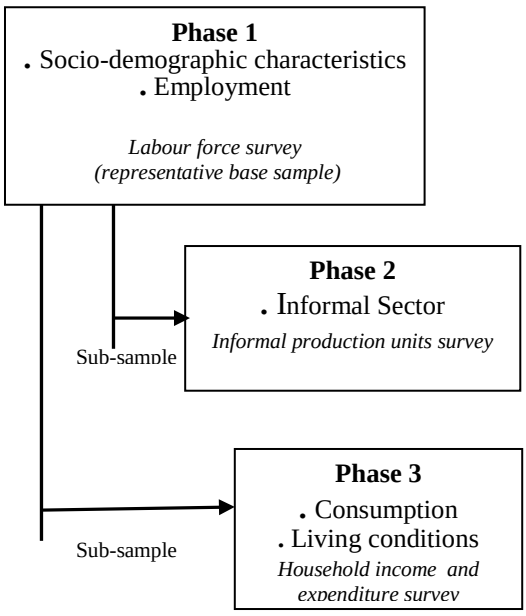
seven principal urban centres. Strengthened by the lessons learnt from this experience, the *1-2-3 survey* has spread to other countries over the last few years. As shown in Figure 6.3, it has been conducted, is in the process of being conducted or planned to be conducted in Africa (Morocco, seven West African countries, Burundi), the Latin America and Caribbean region (El Salvador, Colombia, Peru, Venezuela, Haiti, St. Lucia) and Asia (Bangladesh, China, Sri Lanka, Mongolia, Philippines and Vietnam). The *1-2-3 survey* was designed to act as a generic framework with a flexible structure, which respects a certain number of common characteristics but whose configurations in practice vary according to the needs and specific structure of the existing information systems in the different countries. As a result, in many instances (see Figure 6.3) phase 3 is left out or designed as a stand-alone household income and expenditure survey, and the methodology is reduced to the conventional modular mixed survey conducted in two phases.

3.1.2 Main features

Based on the principle of grafting surveys together, the *1-2-3 survey* is made up of an arrangement of three interlocking surveys, aimed at different statistical observation units: individuals, production units, households (Figure 6.4). The *1-2-3 survey* is an extension of the modular mixed surveys described in section 2. Also, like all other modular mixed surveys that in phase 2 collect information on the number and characteristics of the persons employed in the production units, the *1-2-3 survey* can be considered as a matched employers/employees survey.

The first phase of the *1-2-3 survey* is a labour force survey on employment, unemployment and working conditions of households and individuals (**phase 1: Labour Force Survey**). It allows the labour market functioning to be documented and analysed, and is used as a filter for the second phase, where a representative sample of IPUs (or, more precisely, their heads) is surveyed. Thus, in the second phase of the survey (**phase 2: Informal Sector Survey**), the heads of the IPUs identified in the first phase are interviewed. This phase aims at measuring the main economic and productive characteristics of the production units (production, value added, investment, financing), the major difficulties encountered in developing the business

Figure 6.4: Basic scheme of the *1-2-3 survey*



activity, and the demands for public support by the informal sector entrepreneurs. Finally, in the third phase a specific household income and expenditure survey (HIES) is conducted on a sub-sample of households selected from phase 1. It is designed to estimate the weights of the formal and informal sectors in household consumption by products and type of household (**phase 3: Survey on consumption, formal/informal demand and poverty**). The information collected in phase 3 also allows estimation of households' living standards and monetary poverty, either based on income or expenditures.

In addition, *ad hoc surveys* can be added to this basic architecture, corresponding to varying subject-specific modules and grafted to any of the three phases. This comprehensive approach to measuring the informal sector leads some authors to qualify the *1-2-3 survey* as an example of 'integrated survey' on the informal sector (cf. section 6.3 in Chapter 3).

Each phase is presented below in more detail.

3.1.3 The labour force survey (phase 1): a pillar of the household survey programme

In most countries in the world, primarily the developed countries but also the developing countries of Latin America, Asia, North Africa, etc. – in fact everywhere except for Sub-Saharan Africa – the LFS is at the heart of statistical tools for measuring the economic activities of household members. The LFS can be considered as a generic type of household survey using regular, internationally standardised concepts and methods covering the labour market in general and the working conditions of the population. The LFS is not only the most widespread type of household survey, but also the one around which there has been the largest amount of work on harmonising concepts and methodologies of measurement in order to enable international comparisons, particularly at the instigation of the International Labour Organisation (Hussmanns, Mehran and Verma, 1990).

Two types of arguments plead in favour of using LFSs for monitoring the economic insertion into labour markets of the population in developing countries. Firstly one can mention analytical reasons. The question of employment is central in poor countries, with the vast majority of the population, and especially the most disadvantaged groups, obtaining their income through work, and with institutional transfers of funds (welfare benefits) and income from capital playing only a marginal role. Even more than in other countries, improving the operation of the labour market and access to jobs is central to economic policy. Second and beyond these economic and social considerations, the implementation of the LFSs is justified on technical grounds, involving both the statistical management of surveys and strengthening the institutional capacities of the national statistical offices (NSOs). The LFSs are in fact relatively simple to conduct. The questionnaire is usually short and can be applied in the field in a limited time. In the case of the phase 1 of the *1-2-3 survey*, the short questionnaire of 8 pages takes on average 15 minutes to be completed (Rakotomanana, Ramilison and Roubaud, 2003). This level of performance compares favourably with more complex surveys, where questionnaires sometimes exceed 100 pages, requiring several hours if not days of interviews (cf. LSMS, HIES, etc.). Consequently, obtaining reliable estimators relies upon complex and tiresome procedures. Indeed, this complexity has contributed to the failure of some operations of this kind. By contrast, the per-unit cost of LFSs is limited. Finally, they provide an ideal basis for producing operational stratifications of households, which are relevant for various lines of questioning, and enable further surveys or modules on related subjects to be added.

The operational properties of the LFSs in a context of scarce resources (financial as well as human) combine two significant advantages that should be taken into account:

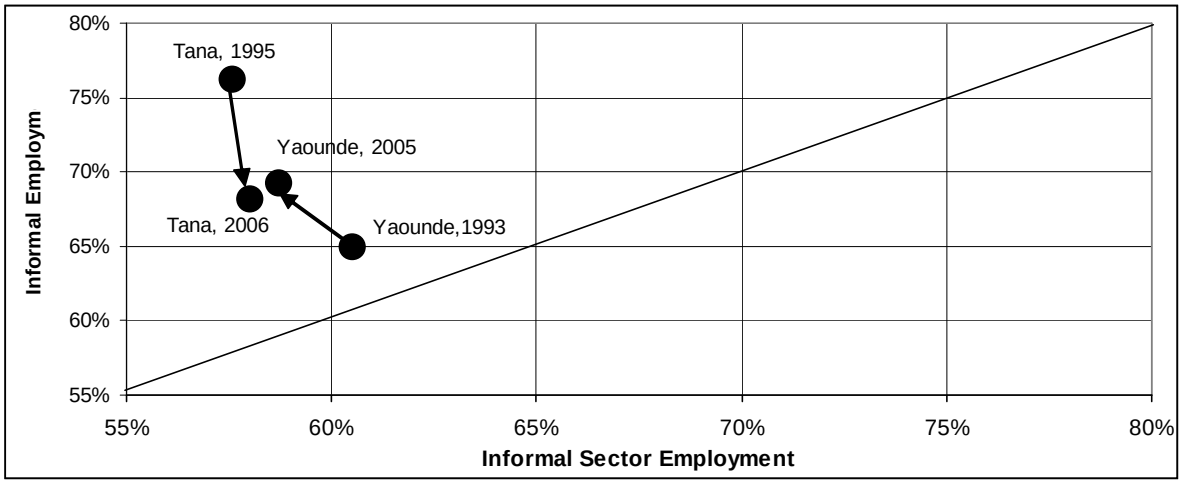
- they enable time series to be build up progressively, which goes right to the foundations of economic analysis. The continuity of sustainable surveys, where both concepts and results are standardised, facilitates the use of (particularly econometric) methods for analysing both micro and macro-economic behaviour;

- they are a useful instrument for motivating teams of statisticians, particularly in poor countries, where household survey systems are still at an embryonic stage. Confronted with immense difficulties, these statisticians need motivating operations with a low risk of failure. As a result, it is advisable to favour surveys that can be controlled by teams of limited size, for which the type of results to be communicated is known in advance and which are not vulnerable to the risks caused by too high funding requirements, which by definition are unpredictable and which cannot be charged to meagre national budgets.

Within the broad category of LFS, the phase 1 of the *1-2-3 survey* has some particular features which shall be underlined. First, the phase 1 has been specifically designed to measure informal sector and employment issues. Concerning the informal sector, a set of questions allows identification, in respect of each employed person, of who belongs to this sector. The questions cover the whole range of criteria admitted by the international definition, i.e. the number of persons employed in the enterprise, different types of registers held (depending on national legislation), and type of accounts (only for independent workers). This information is collected both for the main and the second job. This provides a great deal of flexibility in the operational definition of the informal sector, which can vary according to the purpose of each study (national definition, international comparison, academic objectives). It allows information to be produced on the size of total employment in the informal sector and, using the question on status in employment, on the number of informal production units (or, more precisely, their heads). This last information is of course of crucial relevance to select the representative sample of IPUs for phase 2.

Concerning informal employment, the phase 1 – like any other LFS - represents the most adequate instrument to measure its share (see Chapter 4). Apart from informal sector employment, the questionnaire allows measurement of informal employment in the formal sector using a set of questions about the type of protection linked to jobs: type of labour contracts, payslip, different kinds of allowance (according to national circumstances). Again, the phase 1 questionnaire provides flexibility regarding the criteria of informality to be chosen along international recommendations (see Chapter 2). The example below (Figure 6.5) shows the important changes in informal employment in two cities (Yaounde, Cameroon and Tana, Madagascar) over the last decade. While the share of informal sector employment did not change much, informal employment increased substantially in Yaounde (+4 points), but decreased even more in Tana (-8 points) during the same period. These opposite trends can be explained by different patterns of global integration in a common context of substantial macroeconomic growth. In Madagascar, private formal sector employment has been driven by a rapid expansion of export processing zones (EPZ), which apply international labour standards. Cameroon also registered a strong increase in private formal sector wage jobs by domestic firms, but at the cost of decreasing employment quality and protection.

Figure 6.5: Evolution of Informal Employment by categories in Cameroon and Madagascar 1993-2006



Sources: *1-2-3 Surveys*, phase 1. Madagascar: 1995 & 2006, DIAL/INSTAT. Cameroon: 1993 & 2005, INS/DIAL; author's calculations. Informal employment = Informal sector employment + contributing family workers in the formal sector + wage workers in the formal sector without written contract.

Finally, even if it can be considered as a light survey, the phase 1 questionnaire covers a broad range of information permitting an in-depth analysis of informal and labour market issues: earning functions and returns to human capital (education, experience), on the job training, discrimination (gender, ethnic), segmentation, migration, intergenerational job mobility, job quality (hours, income, bonuses, protection, etc.), job satisfaction, neighbourhood effects, etc. (De Vreyer and Roubaud, forthcoming). To improve the analytical potential of phase 1, a panel component has been introduced in some countries (Benin, Burundi, Cameroon, Madagascar and Peru).

3.1.4 The informal sector survey (phase 2): a supply side perspective

The second phase of the system is carried out among heads of informal production units identified during the first phase (the LFS). Its appeal in the context of monitoring poverty lies in the fact that pockets of poverty are concentrated in the informal sector, especially in urban areas. Paying particular attention to this sector enables a large proportion of the poor to be targeted. Analysing the conditions under which these activities are carried out, economic performance and how informal units are integrated into the productive system, following the input/output table approach, gives a clear picture of the impact of policy on the sector and the strategies of households for which it is a main source of employment or income.

The survey aims to answer precise questions on the role of the sector in the economy, as well as its actual and potential contribution to improving the population's living conditions. As an example, setting up micro-finance systems aims to help the most disadvantaged by encouraging the development of micro and small enterprises. But we have to ask: who benefits from these systems, and what is their impact? On another level, given the limited employment opportunities in the formal sector, does the encouragement of informal sector activities constitute a viable alternative enabling the creation of a growth dynamic? To answer these questions, one needs to understand clearly the economic circuit around the centre of which the informal sector gravitates, by analysing the evolution of its structure and production, the origin of its intermediate consumption, capital, investment, financing and the demand to which it responds.

The standard questionnaire is an individual form. It comprises **7 modules** (12 pages), to which other ones can be added depending on national priorities (for example: social insurance module in West Africa and Vietnam; tax module in Madagascar, etc.):

- module A: Characteristics of the establishment;
- module B: Employment;
- module C: Production;
- module D: Expenditure and costs;
- module E: Customers, suppliers, competitors;
- module F: Capital, investment and financing;
- module G: Problems and prospects.

Previous to these subject-specific modules, the first page of the questionnaire begins with a "Filter module". This module aims at checking that information about the IPUs collected in phase 1 is correct. Relevant information from phase1 on the IPUs of the entrepreneurs selected for the phase 2 (main characteristics of the IPU – address, industry, legal status, type of accounts, registers, type of premises, etc. - and of the holder of the IPU - name, age, gender, relation with household head, status in employment, etc.) is reported *ex ante* in the phase 2 questionnaire. Then, the same information is collected again in the "Filter module". If the answers are consistent, the other modules are applied. Otherwise, the reason of the change between phases 1 and 2 is recorded and, if the selected respondent is not holding an IPU, the survey stops. Apart from delimiting precisely the scope of the phase 2 survey, the "Filter module" can be used to assess the phase 1 data quality (in the same spirit as in post census surveys). Field experience shows that the average time to fill up the phase 2 questionnaire is between 60 and 90 minutes (Rakotomanana, Ravelosoa and Roubaud, 2003; Amegashie et al., 2005).

Compared with informal sector surveys mainly designed to inform informal sector support policies, the phase 2 of the *1-2-3 survey* presents some interesting features. Five main points will be highlighted below. First, the questionnaire has been designed to fit with the suitable national accounts concepts to fuel the elaboration of the main national accounts aggregates. The main challenge is to ensure the compatibility with the daily empirical categories managed by the informal sector entrepreneurs, who in their great majority do not follow formal accounting rules (or even do not hold any kind of accounts). For this purpose, very detailed income and expense tables (product by product) are compiled, leaving to the entrepreneur the choice of the reference period which suits him or her the most for each good and service that he or she produces (from the day to the year, with all combinations in-between). This fastidious procedure is the only one which ensures the reliability of the data collected in the absence of written accounts.

Paragraph 35 (3) of the 15th ICLS resolution acknowledges that production activities of IPUs often overlap with consumption activities of the households of informal sector entrepreneurs. As a result, for some types of expenses (e.g. rent for buildings, fuel for cars, electricity, water) it is difficult to clearly separate business expenses from household expenditure. To avoid that in such cases business expenses are under- or over-reported (and, hence, the value added and the operating surplus of the IPUs are over- or underestimated), the resolution recommends that the expenses concerned should at least be allocated approximately in proportion to their use for business purposes. This can be done, for example, on the basis of information provided by the informal sector entrepreneurs on the percentage of use of the expense for business purposes.

Second, in the perspective of understanding better what is the place of the informal sector in the national economy and its relations with other sectors (public, foreign enterprises, formal domestic enterprises, etc.), for each item listed in the accounting tables, information on the origin of inputs and the destination of the product is collected. In its standard form, the questionnaire distinguishes the following categories: **1.** *public or para-public sector;* **2.** *big trading private enterprise;* **3.** *small trading private enterprise;* **4.** *big non-trading private enterprise*; **5.** *small non-trading enterprise;* **6.** *household/individual*; **7.** *direct exports for destination, or imports for input origin*; **8.** *self-consumption*; **9.** *the IPU (intermediate consumption)*. Furthermore, subcontracting relations are explored in detail. Such information makes it possible to obtain data at the individual level (and *a fortiori* at the aggregate level) to build input/output tables splitting formal and informal sectors both at the product and industry levels.

Third, one important characteristic to deal with is seasonal variations of economic activity in the informal sector over the year. Paragraph 28 of the 15th ICLS resolution recommends that in order to measure seasonal variations in informal sector activities and obtain annual estimates at the aggregate level, data collection should be spread over a period of a whole year by dividing the survey sample into independent subsamples for different quarters or months of the year. However, in some situations such survey arrangement may be too costly and not always feasible because of the required representative geographic spread of each of the subsamples. The phase 2 of the *1-2-3 survey* tries to overcome this problem by including a set of questions about the rhythm of activity over the last year. Once the previous month accounts have been carefully set up, a recall table establishes month by month the level of turnover, distinguishing four options: **0.** *no activity*; **1.** *low*; **2.** *average*; **3.** *high*. To quantify these qualitative appraisals, a specific question captures the estimated minimum and maximum level of turnover (never taken *per se*, but only to estimate the percentage gap between low and high months). This procedure allows, with the highest level of precision possible given, to estimate the annual aggregate values of the IPU at the individual level, which are not directly measurable. Empirical evidence based on African data shows that not taking into account these seasonal factors leads to a substantial underestimation of the mixed income of IPUs (from a minimum of 5% in Senegal up to 39% in Mali; Vescovo, 2007). The procedure is however taxing on the respondent's memory and thus prone to recall errors. Of course the two strategies (with a sample distributed all over the year) can be combined, as it has been done in Mongolia (2007-2008).

Fourth, while the elaboration of national accounts is of prime importance for national accounts and production estimation purposes, this focus is not the only one. The phase 2 also considers other crucial issues, which are more directed at informing public policies. One of the most relevant is the questions exploring in depth the relations between the informal sector and the state: the type of registers and the reason for not being registered, the level of corruption, the compliance with mandatory regulations. This type of question is similar to what can be found in the new generation of business climate surveys to investigate governance issues. Another important topic is about difficulties and demands addressed by informal sector entrepreneurs to public authorities to implement specific policies aimed at facilitating informal sector activities (simplifying registration procedures, scaling up micro-credit structures, targeting training programmes, improving access to equipment, markets and information, adapting tax systems, etc.). These qualitative modules have strategic statistical properties: first they are easier to collect than quantitative production data, so one can ensure that even if the survey failed in its quantitative part, the survey will at least provide valuable information on these issues. Second, field experience shows that usually

informal producers are keener to answer this type of questions than to provide the traditional accounts data. This may also improve the response rate and the quality of answers to the quantitative part, and is a common phenomenon as shown by Herrera, Razafindrakoto and Roubaud (2007) for governance modules in official socio-economic household surveys). It should however be mentioned that information provided by informal sector entrepreneurs on the kind of problems, which they encounter, and the type of assistance, which they need, cannot always be taken at face value. It tends to reflect the problems and assistance, which the entrepreneurs are aware of, but does not necessarily reflect the most important areas for policy intervention to improve their situation.

Fifth, as the employment module (B) lists, for each enterprise, all persons engaged with their personal characteristics (relation with the enterprise head, gender, age, ethnic group, education, training, experience, tenure, etc.), the phase 2 can be considered as a matched enterprise/employee survey. In consequence, it is possible to apply to the informal sector the new developments undertaken with this type of survey, mainly on formal enterprises (Abowd and Kramarz, 1999). In particular, one can reconsider labour market functioning and industrial relations by studying the matching process between employer and employee, taking into account at the same time individual characteristics of employees and enterprises.

Finally, a last important feature of the phase 2 should be mentioned. While impact evaluation protocols are more and more favoured by development actors to assess the efficiency of policies, the phase 2 provides an ideal sampling frame to draw a random sample of the reference group to compare with the treated group (beneficiary of a specific programme or project) on quasi-experimental grounds (Brilleau and Roubaud, 2005). For example, the phase 2 in Madagascar has been used to evaluate the impact of a micro-credit institution targeting the informal sector (Gubert and Roubaud, 2006). A similar survey to phase 2 has been designed and applied to a representative sample of clients of this institution, which have been compared to a comparable sub-sample of phase 2 IPUs and followed over time (three rounds: 2001, 2002 and 2004). This design is suitable to implement all the existing tools used in *ex post* impact evaluation (matching, difference-in-difference techniques, etc.).

Since LFSs were usually not part of the national statistical systems in Sub-Saharan Africa, the African experience with the 1-2-3 survey has been able to overcome the main critics addressed to the modular mixed survey approach: its potential deficiencies in terms of sampling frame. As in all the cases no LFS had been pre-existing in the countries where the *1-2-3 survey* was applied, the phase 1 has been specifically designed to integrate the objective of phases 2 and 3:

- The sample size of phase 1 has been calculated in order to obtain a desired number of IPUs in phase 2 (of course taking into account budget constraints);
- A stratification process has been applied in each phase to optimize the sampling design for informal sector measurement. In phase 1, the EAs have been selected, when possible (as in Madagascar and Benin), along socio-economic strata obtained from the population census, eventually through an intermediate step of developing a master sample of EAs. In all cases, the EAs have been stratified by districts, which made it possible to take into account the unequal spatial distribution of informal sector density. In some cases, an additional stratification variable was added at the second stage of sampling (selection of households within EAs); for example, in Senegal the gender of the household head was used. For phase 2, in all the *1-2-3 surveys* where IPUs derived from phase 1 have been sub-sampled (the majority), stratification of

IPUs has been implemented using phase 1 information. For example, in Madagascar and West Africa, 20 strata were defined by branch of economic activity (10 industries) and the status in employment of the IPU's head (employer and/or own-account worker). The unequal selection probabilities in each stratum have been determined according to the number of IPUs in the LFS sample and the economic potential in terms of development policies. The same type of stratification has been applied for the phase 3 at the household level (gender of household head and household income).

Another strong point of the African experience is the solution given to *the prices issue*. In order to calculate trends in the informal sector's aggregates in real terms, it is not only necessary to replicate the surveys over time, but also to draw up specific price indexes which reflect supply and production structures in this sector. To operate the volume/price breakdown, the various production prices in the informal sector should be computed. In many countries where successive comparable informal sector surveys have been conducted, the informal sector economic aggregates are only available at current prices. To deflate these indicators, two types of price indexes are generally available: the consumer price index (CPI) and the production price indexes (PPI). Nevertheless, none of these two price indexes are reliable measures of price dynamics in the informal sector. On the one hand, the informal production is not destined entirely for consumption and not all consumer goods are marketed by the informal sector. On the other hand, the PPI are generally collected from the formal sector. However, there is no reason to postulate that both the structure of production (for weighting) and the price dynamics are the same in the formal and informal sectors.

The phase 2 of the *1-2-3 survey* provides all the necessary elements to solve this crucial issue, as can be shown in the Madagascan example. Four specific price indexes were drawn up for turn-over, production, value added, and intermediate consumption. The calculation was carried out for nine different industries. Then, the indexes were aggregated to the three main sectors (manufacturing, trade and services) and lastly for the informal sector as a whole. For each industry, the elementary price indexes were calculated at the most disaggregated level of products (four digits), using the unit values collected in the phase 2 questionnaire. These elementary indexes were then weighted by the structures observed in the phase 2 base year (1995 in Madagascar), taking account of both the weight of products within each industry and their origin or sector destination (formal or informal). The price of a product for intermediate consumption varies, for instance, depending on whether it has been purchased from a supermarket or from a small informal business. The value added price, which cannot be observed from the market, was deduced from production and intermediate consumption prices.

3.1.5 The survey on consumption (phase 3): a demand side perspective

Phase 3 of the *1-2-3 survey* is basically a household income and expenditure survey (HIES), conducted on a sub-sample of households surveyed in phase 1. Classically, it aims to determine the level and structure of household consumption, but more originally, the survey has been designed to estimate the share of informal sector in household consumption (and household fixed capital formation). Two methodological points are worth stressing: as a HIES, phase 3 is conceived to build on the experience accumulated in this field of research; in addition, it serves to measure consumption items purchased in the informal sector.

Following the example of other surveys of the same kind, phase 3 provides estimates of household consumption and incomes. It enables poverty lines to be calculated and the traditional indicators of monetary poverty to be estimated (incidence, intensity etc.). This is of strategic importance as poverty reduction has become the main objective of development policies in developing countries (Cling, Razafindrakoto and Roubaud, 2003). From a methodological viewpoint, the survey is confronted with the usual trade-off between reliable estimators and rapid communication of results. The option chosen was to position phase 3 halfway between the traditional HIES and the LSMS survey. The use of accounts diaries (for daily expenses) covering two weeks provides a better measure of consumption and thus of poverty than is given by surveys, which are based solely on retrospective questioning. The fact that the survey takes less account of seasonality than traditional HIES, where data are collected throughout a full year, can be explained by the in most cases urban coverage of the survey, and is partially compensated for by shorter time spans for delivering the information. Nevertheless, the phase 3 standard sample is divided in two successive waves of 15 days each, covering a full month to take into account intra-month variations (for example, in many countries civil servants or private wage workers receive their pay on a fixed day of the month, often at the end of the month). Beyond daily expenses, which are only recorded to estimate food consumption, retrospectives modules are used to collect data for all other consumption categories (health, education, clothing, etc.). The reference period for these modules varies from 6 months to 2 years, according to the frequency of purchase.

The standard questionnaire is a household form. It comprises **15 modules**, to which other ones can be added depending on national priorities:

- 1 module for the diary recording household expenditures during 15 days;
- 1 retrospective module (over 12 months) for important or exceptional expenditures of cereals and other food products;
- 1 retrospective module (over 12 months) for expenditures realized during celebrations, funerals and other ceremonies;
- 9 retrospective modules for non-food expenditures (final consumption), with variable lengths (over 6 or 12 months);
- 1 retrospective module (over 12 months) for taxes;
- 1 retrospective module (over 24 months) for construction expenditures (own dwelling);
- 1 retrospective module (over 6 months) on monetary transfers from and to other households.

From an analytical viewpoint, the main purpose of phase 3 lies in its estimates of amounts spent by each household, keeping track of where products were purchased (or obtained for self-consumption), and in particular their origins in the formal or informal sector. For each product, apart from collecting information on the type of product, the quantity, the unit price, the total value, and the place where the product has been bought (or obtained) is captured. The standard typology of place of purchase is:
INFORMAL: **1.** *Self- production;* **2.** *Hawking, street;* **3.** *Seller's home, small informal shop;* **4.** *Market;* **5.** *Other informal place of purchase.*
FORMAL: **6.** *Supermarket;* **7.** *Formal shops and stores;* **8.** *Public sector (friendship store, etc.);* **9.** *Other formal place of purchase.*

Of course, drawing a perfect line between the formal and informal sector is out of hand for the interviewees. In some cases, the respondent within the household does not precisely know

if the supplier of a product keeps accounts, has registers, or what is the number of persons working in the enterprise. However, these cases are limited as the task is mainly to distinguish between categories 5 and 9. If there is no way to have a decisive clear cut between these two categories, errors can be adequately reduced. First, the typology of place of purchase can be refined (up to 39 items in Colombia, 2001, and 42 in Morocco, 2000). Second, specific instructions can be given in the field: interviewees are informed *ex ante* of what is the exact definition of the informal sector, so that they can provide valid information about the characteristics of their suppliers (and even, for accounts diaries, collect themselves additional information about the formality status of their suppliers). In this context, we have to take into account the usual consumption patterns: in general, consumers keep the same suppliers for long periods and have some kind of personal proximity with them (especially in the informal sector). Potential errors can also be tracked and corrected *ex post*, elaborating on the interviewer's knowledge of the characteristics of suppliers (by industry, neighbourhoods, etc). Finally, checks and corrections can be undertaken during the data processing step. The information given by the phase 2 about the structure and characteristics of the informal sector is of particular relevance here. The last step consists of checking the consistency of estimates (at the aggregate and more detailed levels) between household consumption purchased in the informal sector given by the phase 2 (destination of outputs) and by the phase 3.

The consolidation process of the phases 2 and 3 allows detailed resources / uses balances by products to be elaborated: Production = Intermediate Consumption (differentiating different industries and sectors, see destination codes of phase 2) + Capital Formation + Changes in Inventories + Household Final Consumption (differentiating different types of households) + Exports; see Chapter 9), isolating informal sector goods and services. This is of crucial relevance for building national accounts, informal sector satellite accounts, and social accounting matrices.

Additionally, for each product two supplementary questions are asked. First, the phase 3 collects information on the reason for which each supplier has been chosen, distinguishing (in the standard version) between: **1.** *less expensive;* **2.** *better quality;* ***3.*** *access to credit, payment facilities;* **4.** *proximity;* **5.** *Good relations;* **6.** *does not find elsewhere;* **7.** *other reason*. This question makes it possible to determine the behaviours of different categories of households (according to their wealth, sources of income, etc.) in their decisions to buy a product in the formal or the informal sector. Second, the respondent is asked to specify the country of origin of each product consumed. Here again, some cases raise the issue of reliability of the answer provided. To improve data quality, the same type of strategy as in the case of the place of purchase has been put in place. This question is of particular relevance in countries where trade statistics are not reliable and where important flows of imports enter the country "informally" (without being registered by customs procedures). This is the case of many developing countries (and even developed countries affected by smuggling activities in respect of particular products such as cigarettes, etc.

3.2 The UN Development Account Project 'Interregional cooperation on the measurement of informal sector and informal employment'

This section presents the main features of nationally/territorially-representative surveys conducted by Sri Lanka, Philippines, Mongolia, St. Lucia and the West Bank and Gaza Strip

that represent variations of the modular mixed survey design described in this chapter. The section concludes with a discussion of issues and challenges in their implementation.

The surveys go only as far as the first two phases of the '1-2-3' surveys described in section 3.1.; they are thus referred to as '1-2' surveys. These surveys were a core component of the 2006-09 United Nations Development Account Project "Interregional Cooperation on the Measurement of the Informal Sector and Informal Employment"[4]. This project aimed to contribute to the global efforts at identifying, adapting, and testing cost-effective and sustainable data collection strategies for estimating employment in the informal sector, informal employment and the contribution of informal sector enterprises to the economy as measured by GDP.

3.2.1 '1-2' Survey methodology of the Development Account Project

In the '1-2' survey methodology implemented under the Development Account project, phase 1 is based on an existing household labour force survey (LFS). For phase 2, the project methodology broadened the scope of data collection to household unincorporated enterprises with at least some market production (HUEMs). A HUEM is a household production unit defined as follows:

HUEM0. *Ownership.* A HUEM belongs to the *household* sector.
HUEM1. *Legal organization*. A HUEM is not constituted as a separate legal entity independent of its owner(s).
HUEM2. *Book-keeping practice.* A HUEM does not typically keep formal (written) accounts.
HUEM3. *Product destination*. A HUEM is different from other household enterprises in that it should sell (or barter) at least part of its production on the market.

Following the definition of informal sector enterprises set out in the 15th ICLS resolution on employment in the informal sector, *informal sector enterprises* are a subset of HUEMs that are delineated from 'formal' HUEMs by applying the criteria of *non-registration* and *employment size*. In the '1-2' survey method implemented by the project, production units associated with jobs of employed persons are classified into mutually exclusive categories by institutional sector (see Figure 6.6). In particular, production units within the household sector are classified into HUEMs and non-HUEMs that include households producing for own final use. HUEMs are then further classified into informal sector enterprises and formal sector enterprises within the household sector based on country-specific criteria concerning non-registration and employment size. HUEMs so identified are the statistical units for the phase 2 survey.

3.2.1.1 Phase 1 of '1-2' survey

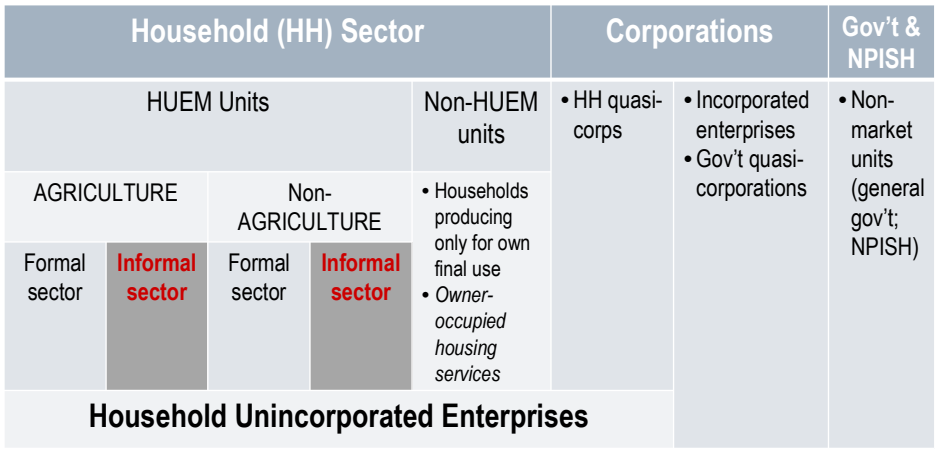
The first phase of the '1-2' survey implemented by project countries/territories had the following common design elements:

(a) Phase 1 survey was based on an existing LFS with data collection on either a quarterly or a monthly basis (to take into account seasonal variations of informal sector activities). The

[4] Two additional territories/countries, i.e. the West Bank and Gaza Strip and St. Lucia, participated in the project and conducted '1-2' surveys following the same methodology. For more information see: http://www.unescap.org/stat/isie/index.asp

existing LFS questionnaire and survey operations were modified to collect data to address three objectives:

Figure 6.6: Production units in the system of national accounts



i. Estimating employment in informal sector enterprises
ii. Estimating informal employment[5] (including employment in informal sector enterprises, as in (i))
iii. Constructing a sampling frame for the phase 2 statistical units; i.e., HUEMs

(b) The modification to the existing LFS questionnaire[6] involved the inclusion of additional questions designed to meet the above-mentioned objectives by obtaining information on both main and secondary (if these exist) jobs of employed persons. These were questions to identify HUEMs among the production units where jobs are undertaken and on non-registration and employment size for identifying informal sector enterprises from among the HUEMs as well as questions for determining informality of jobs of employees. These questions were either integrated into the existing LFS questionnaire (Phase 1 integrated questionnaire) or in a separate questionnaire developed for the purpose (Phase 1 modular questionnaire). A generic questionnaire that was adapted by project countries is shown in Annex 1.A[7].

(c) Countries utilizing the modular questionnaire modified the survey operations of the existing LFS to complete the interview for phase 1. In these cases, the LFS interviews for the household members were first completed, and afterwards the phase 1 questionnaire was administered to all employed persons in the household as identified in the LFS interview.

(d) The sample design for the phase 1 survey followed that of the LFS, which for all participating countries/territories is a two-stage design. Stage 1 sampling units are unit areas

[5] The definition of informal employment used here is in accordance with the 17th ICLS guidelines as described in Chapter 2 of this manual.
[6] The LFS questionnaires of the project countries/territories followed international recommendations on labour force statistics in identifying employed persons. However, not all countries collected information on secondary jobs and this was introduced in the phase 1 questionnaire.
[7] The generic phase 1 questionnaire developed by the project was essentially based on the questionnaire used in the continuous Labour Force Survey of the Republic of Moldova (see Hussmanns, ILO Working Paper #53 *Measuring the informal economy: From employment in the informal sector to informal employment*, 2004) as discussed in Chapter 4 of this manual.

and stage 2 sampling units are households or dwellings. Thus, the base sampling weights for estimates generated from the phase 1 survey are the final sampling weights of the LFS survey.

3.2.1.2 Phase 2 of the '1-2' survey

The phase 2 surveys were nationally/territorially representative surveys of household unincorporated enterprises with at least some market production (HUEMs) and had the following common design elements:

(a) The phase 2 survey aimed to obtain data on informal sector enterprises for:
   i. Estimating production, expenditures and value-added of informal sector enterprises and their contribution to total GDP
   ii. Estimating employment and compensation in informal sector enterprises
   iii. Generating information on organization, capital formation, business environment, problems and prospects, and policy-related issues

(b) The sample frame of HUEMs was built up from information collected in the phase 1 survey that identified HUEMs associated with jobs (main and secondary) of employed persons with status in employment of employer or own-account worker. The surveys in the West Bank and Gaza Strip and the Philippines covered both agricultural and non-agricultural HUEMs while the other three countries covered only non-agricultural HUEMs.

(c) All except the survey in the Philippines, which selected a probability sample[8], enumerated all HUEMs ('take-all' sample) in phase 2. Thus, phase 2 base weights for the 'take-all' sample HUEMs were the phase 1 final weights of the corresponding employed persons; for the survey in the Philippines, the phase 1 final weight was multiplied by the inverse of the probability of selection of a HUEM.

(d) Enumeration of sample HUEMs for phase 2 was conducted after phase 1, either immediately following the interview for phase 1 (no time lag) or with a short (e.g., two weeks for Sri Lanka; one month for the Philippines) time lag between phases.

(e) A model questionnaire (see Annex 1.B), based on the phase 2 questionnaire of the '1-2-3' surveys described in Rakotomanana[9] et al (2003), was developed by the project and adapted by countries/territories.

3.2.1.2 Country-specific implementation strategies

The main design elements of the '1-2' surveys implemented by the project are listed in Figure 6.7. The specific implementation strategies of the five national/territorial surveys described in terms of these correspond to four models as summarized in Figure 6.8.

[8] The sample design sub-sampled PSUs from the phase 1 sample; all HUEMs identified in phase 1 for these PSUs were enumerated. In sub-sampling PSUs, a stratification scheme was applied to ensure adequate representation of all branches of economic activity.

[9] Rakotomanana F., Ravelosoa R. and F. Roubaud, *The 1-2-3 survey of the informal sector and the satisfaction of household needs in the Antananarivo conurbation*, 2003. InterStat No. 27, September, 59-88.

Figure 6.7: Key survey design elements of '1-2' survey

Element		Description
A	**Questionnaire design for phase 1**	*Additional questions for identifying HUEMs, IS enterprises and informal employment*
A1	Integrated questionnaire	Incorporated in LFS questionnaire
A2	Modular questionnaire	Independent questionnaire
B	**Survey operations for phase 1**	*Procedures for phase 1 interviews*
B1	Integrated approach	With integrated questionnaire, data collection for phase 1 is part of LFS interview
B2	Modular approach	Modular questionnaire is administered *after* the interviews for the LFS are completed
C	**Time lag between phase 1 and phase 2 interviews**	*Time lapse between phase 1 and phase 2 interviews*
C1	No time lag	HUEM survey questionnaire is completed right after phase 1 interview
C2	Some time lag	Survey operations for phase 2 are independent of phase 1; enumeration is done after phase 1 survey
D	**Sampling frame: industry coverage**	*Coverage of economic activity sectors*
D1	All activities	All HUEMs are included in sampling frame
D2	Non-agricultural activities only	Only HUEMs whose primary activity is in non-agricultural sectors are included in sampling frame
E	**Sample design for phase 2 survey**	*Selection of sample HUEMs for phase 2 survey*
E1	Complete enumeration	All HUEMs in the sampling frame are enumerated
E2	Sample enumeration	Sample of HUEMs is selected (Philippines only): subsample of PSUs of phase 1 is selected; all HUEMs identified in phase 1in the subsample of PSUs are enumerated

Figure 6.8: Implementation strategies for '1-2' survey: Five countries/territories

Model	Design Elements	Country/Territory
1	A1, B1, C1, D1, E1	West Bank & Gaza Strip, St. Lucia
2	A1, B1, C1, D2, E1	Mongolia
3	A2, B2, C2, D2, E1	Sri Lanka
4	A2, B2, C2, D1, E2	Philippines

3.2.2 Issues and challenges[10]

3.2.2.1 HUEM as statistical unit for data collection

The '1-2' surveys used the HUEM as a statistical unit. As a result, analyses concerning informal sector enterprises can be done on these data, using various cut-offs for employment size or types of non-registration in defining informal sector enterprises. By not pre-selecting

[10] The observations made in this subsection are based on the preliminary reports and initial assessments of experiences of project countries/territories available in December 2009.

informal sector enterprises according to national definitions as the statistical units of the survey and including questions asked uniformly in all project surveys, it is possible to use a definition that is consistent across the surveys.

In addition, the '1-2' surveys provided data for identifying households producing for own final use which is an element of the definition of informal employment. A suggestion is to improve the understanding of the question:

QHUEM3 1: Does the enterprise you own/where you work sell or barter its goods and/or services?

1. Yes, at least some part of it on a regular basis	
2. Yes, at least some part of it from time-to-time	
3. No	*Not HUEM*
4. Don't know	*Not HUEM*

to mean that products are intended to be sold/bartered in the market on a *regular* basis.

3.2.2.2 LFS and phase 1: integrated or modular?

Estimation of employment in the informal sector and informal employment should be considered objectives of a standard labour force survey. It is thus recommended that the phase 1 survey be completely integrated with the LFS. This was the case in fact for Mongolia, the West Bank and Gaza Strip, and Saint Lucia. However, these changes to existing systems may result in unforeseen delays in the statistical release calendar for the LFS. This was the reason why Sri Lanka and the Philippines opted for the modular approach to data collection.

As shown by the experience of Sri Lanka, the modular approach increases the chances of inconsistencies in responses to similar questions in both questionnaires and of unit non-response in phase 1, i.e., persons identified as employed in the LFS are not interviewed in phase 1.

3.2.2.3 Dependency between phase 1 and phase 2: sample size and sample design

Phase 2 is dependent on phase 1 in that phase 1 generates the sample for the HUEMs to be surveyed in phase 2. Consequently, the number of sample HUEMs is dependent on the number of sample households of the LFS survey and small LFS samples means small phase 2 sample sizes. As a result, not all branches of economic activity could be suitably represented for estimating value-added of informal sector enterprises at the desired levels of disaggregation. For example, the Mongolia survey results indicated that informal mining and transport, both known as important segments of the informal economy, were undercovered. Anticipating this, the surveys of the West Bank and Gaza Strip and Saint Lucia expanded the sample sizes for phase 1. In general, to obtain reliable and direct estimates on household enterprise statistics disaggregated by industry, size, geographic location, type of ownership and other characteristics, the size of the sample in the first phase should be large enough.

A second dependency is that the sample design of the LFS is based on the distribution of the population and not on the distribution of enterprises. Hence, the '1-2' survey might produce less efficient estimates than dedicated informal sector surveys or any other type of survey for

which the sample design is drawn from a distribution of the enterprises, formal or informal[11]. A proposed solution - post-stratification using auxiliary information on the total number and distribution of household enterprises or informal sector employment available from economic censuses to adjust marginal distributions of the estimates to the distribution of HUEMs in the population (followed by appropriate adjustments in weighting) - could not be implemented as none of the project countries had conducted an economic census.

With a sample design involving sub-sampling of LFS PSUs for phase 2, large variations in the weights for phase 2 of the Philippines '1-2' survey were observed. Thus, trimming of weights was recommended and implemented.

3.2.2.4 Time lag between phases

As a rule, the shorter the lag in the data collection between the first phase and the second phase surveys, the higher are the chances that the HUEMs can be traced and interviewed in the second phase. With a two-week lag between phases for the Sri Lanka '1-2' survey 87 percent of identified HUEMs were located and interviewed, 6 percent were non-responding units, 2 percent were identified as HUEMs in phase 1 but were no longer operating, and 5 percent were identified as HUEMs in phase 1 but were determined to be non-HUEMs based on the phase 2 questionnaire. Non-response adjustment factors were applied to account for non-responding units.

3.2.2.5 Data quality

The surveys confirmed difficulties in obtaining information of good quality about enterprise characteristics from employees. Employees do not generally know the characteristics of the enterprise where they work especially legal status, accounting practices, and registration with the public authorities. Thus, only HUEMs associated with self-employed persons were included in the frame.

For the phase 2 survey, it is known that the quality of the information is lower in the case of proxy respondents. Thus, the phase 2 respondent should be the owner or operator of the HUEM. This became an issue in the typical case where the LFS respondent was not necessarily the employed person. Thus in '1-2' surveys where phase 2 immediately followed phase 1 interviews, provisions needed to be made to ensure that the employed person himself/herself was the HUEM survey respondent. For one, the interview should take place in the premises of the HUEM, if the premise is fixed.

The questions on the identifying criteria of HUEMs and informal sector enterprises are repeated in the second phase questionnaire. This practice verifies the accuracy of the information, which may have been provided by a proxy respondent during the LFS. Indeed, there were cases where these were not consistent and suitable editing rules were applied.

Due to the absence of written accounts, the HUEM questionnaire was designed to enable the respondent to reconstruct this information as accurately as possible. For example, data on the value of production and sales by specific activity, and the value of raw materials by product, was collected for a short reference period (selected by the respondent), say a week. To obtain

[11] Verma, Vijay. *Sample Design Considerations for Informal Sector Surveys*, 1999. Proceedings of the Conference of the International Statistical Institute, Istanbul, available at http://www.stat.fi/isi99/proceedings.html

accurate weekly values, a worksheet format was used to record data on unit, quantity, and unit price for the selected period (see Annex 1.B).

To estimate annual sales/output values, the questionnaire design took into account seasonal variations in the level of operations of the enterprise. For example, along with data collected for the past month, questions on the intensity of the business activity during each month of the year and on the average level of receipts/profits in the months of high/low business activity was also collected. The past month values were used to generate annual values using these factors.

However, the project experiences revealed weaknesses in developing and applying data editing routines as well as imputation procedures for item non-response for the economic variables.

3.2.2.6 "Sustainability" in the context of value-added estimation for informal sector enterprises

The project results show that there are no major impediments to enhancing existing LFS questionnaires so that they meet requirements for estimating employment in the informal sector and informal employment. That is, a regular phase 1 survey can (and should) be readily fielded by countries.

On the other hand, how often a phase 2 survey should be conducted for the purpose of estimating informal sector GDP is a more difficult question to answer. The project results demonstrated that the phase 2 survey can provide direct estimates and if so should be conducted as often as required for national accounts estimation. At the very least, as an enterprise survey it should have its place in the survey cycles of existing enterprise/establishment surveys.

3.3 The National Micro Business Survey in Mexico (ENAMIN)

Mexico's experience with mixed modular surveys on the informal sector goes back to 1987 with the joint research programmes of the French Institute for Research on Development (IRD) and the Mexican National Institute of Statistics and Geography (INEGI), specifically the Pilot Survey on Informal Sector (EPSI by its Spanish acronym) and in 1988/89 the National Survey on Informal Economy (ENEI) which covered Mexico's seven largest urban areas. In 1992 the National Survey on Micro Businesses (ENAMIN) was begun by INEGI with the financial support of Mexico's Ministry of Labour. The objective of this survey was the collection of data on a regular basis on both formal and informal non-farm businesses as a regular module of the Labour Force Survey (LFS). Initially data were disseminated on all micro businesses and later specifically on the informal sector, as a basis for estimating its share of the GDP.

The operational definition of a *micro business* encompasses all non-farm activities - without prejudging if they are formal or informal, - with the following size criteria:

- for trade, transport, construction and services no more than 6 persons (including the head of the enterprise);

- for manufacturing a criterion of up to 16 (including the "head") was set so as to complement the manufacturing establishment survey; the sample of the latter covered most of gross production, and consisted mainly of medium and large economic units.

The (first phase) questionnaire of the ENEI contained elements to identify employment in the informal sector. The conceptual rigor of the questionnaire design was improved further when ENOE replaced ENE as the Mexican LFS (see Chapter 4) in 2005. The second phase (ENAMIN) was addressed to heads of non-farm micro-production units, including professionals, operating with or without fixed premises as a main or a complementary activity. The design of these surveys reflects certain analytical objectives, specifically to contrast the economic performance of formal and informal units, the social backgrounds of those leading informal and formal activities as well as elements that contribute to understanding the rationality and comparative advantages and disadvantages of formal and informal units. These data provide the basis for testing a wide number of prevailing theoretical approaches to the informal sector, for example does the informal sector arise out of the need to avoid entry barriers, cumbersome bureaucratic obstacles and transition costs (De Soto, 1987); is the informal sector a last survival resource; is it due to deep segmentation of the labour market or do some unintended incentives contribute to the phenomenon; such as a voluntary choice above anything else (Maloney, 1998)? Unfortunately the possible uses of ENAMIN to answer these questions have not yet been fully explored by researchers or academics.

An important concern in Mexico addressed by the ENAMIN data relates to the potential tax contribution of informal production units. Many in Mexico have speculated for some time that the rather low proportion of income taxes in the GDP is due to a narrow base of tax registered contributors, with the informal sector as one of the main tax evasion suspects. ENAMIN data has deflated that view. Moreover ENAMIN came up that, as an additional finding not taken into account in fiscal policy, informal production units face uncertainty as well as catastrophic risks –either personal or familial. Some policy makers have understood that before discussing the fiscal implications of tax evasion by the informal sector, it is important to solve issues such as access to a health care scheme. This understanding helped to launch the so-called *"Seguro Popular"* in 2003 as an alternative to the health care services of the social security schemes which covered only salary workers having a typical employee-employer contract. This in turn has generated a wider debate with some proponents pointing out that a dual scheme will not work and that the best public policy is movement toward a universal health care scheme (Levy, 2008). In any case the data provided by ENOE (first phase) and ENAMIN (second phase) on both informal employment in its wide sense as well as specifically on informal sector have had important consequences in shaping policy making and the debate surround it.

The objective of ENAMIN is to provide data on the informal sector, its relation with State authorities, public programs, the economic environment and other information relevant for policy making and pertinent to the measurement of its economic contribution.. ENAMIN is not designed as a source of data on employment in the informal sector because employment data are collected in ENOE, the Mexican LFS or first phase. ENOE supplies these basic figures, on a quarterly basis, alongside those of unemployment, underemployment and non-protected salary workers (those without non pecuniary benefits or health care services). This information is now available from 1995 onwards.

It is important to discuss why Mexico adopted the first two stages of the 1-2-3 method without taking on the third. One reason concerns the history of Mexico's statistical system. When the first trials of the mixed modular approach were made, Mexico had well developed household surveys with no major statistical gaps. In fact the oldest household survey dating back to the mid-1950s is the income-expenditure survey; first designed to provide the commodity weights needed to estimate the Consumer Price Index (CPI) and later incorporating additional contents in order to allow measures of poverty and well being. ENIGH – the income-expenditure survey – was firmly established in the Mexican statistical system and it would not be easy to add another expenditure type survey even if it pursues more specific aims.

In addition there were doubts that the LFS sample which provided the sample for the second phase module was suitable for estimates relating to a range of expenditure issues. Even to provide detailed estimates of informal sector activities by industry is a challenge for the modular approach.

Operational considerations were also important: the Mexican LFS is a semi-panel continuous survey, that is, each group of selected dwellings remains in the sample for five quarters before being replaced by a new group. There was concern that to add many questions in a given quarter may put at risk the acceptance of the next interview, thereby increasing the total non-response rate in the future to levels beyond those assumed by the LFS sample designers. In the case of Mexico this is a particularly sensitive issue considering that the LFS generates estimates for many sample domains: national (distinguishing four types of population densities), 32 states (provinces) and main urban areas. However if Mexico's LFS were conducted only once a year or every two years the 1-2-3 strategy may not pose such operational problems.

Mexico's estimates of the share of the informal sector in the GDP rely mainly on a supply side approach rather than on the demand side because there is no phase 3 survey. Since the data for the orthodox approach are not available, national accountants in Mexico have opted to disseminate data on the significance of the informal sector in the GDP by means of a Satellite Account *("Cuenta Satélite del Subsector Informal de los Hogares").*[12]

Another limitation to take into account is that from the ENAMIN first wave in 1992 to 2002 the sample (actually a subsample of heads of micro businesses taken from the LFS or first phase) was restricted to urban areas. For the rest of the country the coefficients of mixed income per capita and wages were applied respectively to heads of businesses and paid employees in informal production units in areas within the reach of the LFS but out of the ENAMIN sample. This is a big assumption: namely that productivity in urban areas – with the advantages of urban economies of scale - is the same as in the rest of the country. As a consequence, this may result in an overestimation of the informal share of the GDP. However in practice a supply-side approach to the value-added generated from ENAMINs 1992-2002 data might be an underestimation because informants lack incentives to be truthful with regard to their actual income, sales or production levels. Generalizing these estimates to the rest of the country - that is to informal production units in the less productive non-urban areas - may compensate to some extent for the tendency to underestimate in the supply side

[12] A possible solution might be to incorporate in ENIGH (the household income-expenditure survey) questions on the characteristics of the supplier of the goods and services purchased to determine as a proxy if they belong to the informal sector. However to do that for each expenditure item may overburden the already most demanding and time consuming survey for the respondent which INEGI has ever devised.

approach. However this approach (generalization of coefficients) is invalid and never has been used in Mexico to assess the socioeconomic conditions of those outside the high density urban areas. The assumptions introduced for national accounts calculations were made so as not to affect other kinds of analysis that could be conducted using the micro-data.

ENAMIN 2008 overcomes the limitations of its predecessor because for the first time its sample is the total from the first phase, including all owners of informal production unit identified by the LFS in urban, intermediate and rural areas. The weighting factors from both ENOE (LFS) and ENAMIN are basically the same except to the extent that ENAMIN 2008 has its own non-response rate due to attrition (see section 2.2.1 above), its weighting factors are additionally adjusted in order to compensate for a loss of observations from the first phase.

Given the experience in Mexico with second phase modules on other topics where the LFS is the first phase, the first concern of both operational and analytic teams associated with ENAMIN was the large proportion of lost observations that occurred when information collected in the LFS was not corroborated by ENAMIN. With other survey modules, if there was a mismatch, confidence was placed on the second phase module if the information was provided by a respondent who was directly concerned with the questions addressed (self-response). However with ENAMIN there were grounds to believe that information provided by proxy respondents in the first phase (LFS) was not necessarily unreliable because the micro entrepreneur, in many cases, may have good reasons to hide information from any institution tracking her/his activities.

Figure 6.9 shows the sample size selected for each ENAMIN wave from 1992 to 2008 (second column); the non-response rates affecting the second phase due to various causes (third and fourth columns); the percentage of the sample where the identification of heads of micro-business in the first phase (LFS) was not corroborated by ENAMIN (fifth column); and the percentage of the sample with full and effective interviews (sixth column).

From the beginning of ENOE in 2005, questions on the nature of the activity (see Chapter 4) were collected in the first phase. For example, data were collected not only on status in employment, industry and size but also on registration, type of accounts and the use of the services of accountants. These data items became elements to validate information on businesses collected in the second phase. This in turn sparked a debate among members of the analytical team on whether it was better to stick to the orthodox approach for the 2008 survey making use of ENAMIN's specific set of weighting factors adjusted by the loss of observations, or rather ignore them and maintain the original LFS weighting factors of any record, including those corresponding to the observations lost in ENAMIN. The latter approach would attribute the same value-added coefficients of a similar group of production units that in ENAMIN were effectively interviewed, and share all key features already identified in the LFS, including sample strata. The second approach was adopted because the traditional weighting factor approach to adjustment tends to ignore the fact that the observations missed are concentrated in some part of the spectrum rather than evenly distributed. The point of including all missing records with their original weighting factor was not to depend on that assumption. Trial exercises, which compared the total value-added obtained from the records effectively interviewed by ENAMIN and expanded by means of weighting factors adjusted by the total non-response specific to ENAMIN, were significantly smaller than the total sum obtained from those records plus those missed, all with their original LFS weighting factors (that is with no further total non-response adjustment). This

suggests that those missed in the second phase are predominately respondents having economic reasons to avoid ENAMIN exploration.

Figure 6.9: Information on ENAMIN samples 1992-2008

Year	Sample size: eligible records identified in the LFS to be visited by ENAMIN	Non-response because individuals cannot be found or a contact cannot be made after two weeks	Non-response due to refusal of the second phase interview (ENAMIN) by the self-respondent	Cases where the self-respondent denied he/she conducted a micro-business or gave information not matching the information collected in the first phase (LFS)	Second phase interview accepted and matching with the information provided in the LFS	Number of urban areas in the ENAMIN sample	Fieldwork period for ENAMIN
1992	13 363	9.8%	4.5%	18.1%	67.6%	16	22/01-30/04 1992
1994	12 243	5.0%	6.6%	4.0%	84.4%	16	31/01-14/05 1994
1996	13 219	6.1%	3.0%	5.7%	85.3%	16	31/01-14/05 1996
1998	15 971	6.1%	6.0%	12.7%	75.2%	44	02/11-1998 01/03 1999
2002	15 000	5.2%	4.4%	15.0%	75.4%	45	11/02-15/06 2002
2008	30 063	3.5%	6.7%	7.2%	82.5%	National	27/11-2008 25/01-2009

Source: INEGI. National Labor Force Survey (1992-2008); Encuesta Nacional de Micronegocios (1992-2008).

The debate however is still open because the value added of those missed in ENAMIN but included in the LFS was the result of an imputation taking as reference the most similar records effectively interviewed by ENAMIN. Hence the difference is not altogether implausible given the orthodox procedure which is blind to these similarities. The difference may have to do with the imputation technique. Indeed a real challenge that this type of survey faces is how to deal with both total and partial non response. This will be taken up at the end of the section.

The ENAMIN questionnaire covers fifteen topics:

1. Verification of the information supplied by the first phase
2. Migration condition
3. Labor background
4. Business beginnings (how, when and why)
5. Types of accounts kept/business registration
6. Premises
7. Hours worked
8. Data on employees and others helping in the business
9. Equipment/physical assets/investment
10. Problems faced in conducting the business/activity

11. Unexpected contingencies/business expenditures and incomes/sales
12. Loans and financing
13. Formal training (in order to improve business performance)
14. Other support programs
15. Prospects/expectations

Any eligible individual does not need to answer all questions or even all the generic topics; the questions that need to be answered depend on the sequence followed (the structure of the ENAMIN questionnaire is shown in Annex 2). The interview extends for about an hour and eight minutes and is held either at the premises - if the business has one - or in the household. The interviewers ensure that the self-respondent was questioned.[13]

As the diagram shows, the placement of topics is related to the economic flow of activity trisected by industry: one section focuses on manufacturing, another on trade and the third on construction, transport and services. If an individual has more than one business located in different places, the instruction given to the field staff during training specifies that if the business is in the same industry with the same type of goods or services provided, all locations are to be considered as a single economic unit.

Strategic as well as tactical decisions were made concerning the questionnaire design. Considering the mistrust many entrepreneurs in informal production units have about giving information on their economic activities, it is important to lay the groundwork before taking on the most sensitive topics in the interview. It is also important to gather relevant information using more than one route of exploration rather than to place all eggs in one basket. For example, in addition to reconstructing by means of the questionnaire a simple production account of the business, information should also be obtained on the mark-up (the difference between market prices and costs) of the main products or services sold and on economic flows. Information which allows for cross checking needs to be included at different places in the interview.

The declaration of income should not be taken at face value. What the respondent says he/she earns working independently is averaged with the income he/she thinks would be made if working for salary – a question which is also asked. The first declaration (the straightforward response on the allegedly actual incomes obtained from the business) is taken as the floor, the second one as the ceiling. Once averaged, this exploration on what national accountants identify as mixed income is called Approach 1. A tactical design feature is to ask first what level of income as a salary worker would be acceptable before asking about actual incomes in conducting the business. Once a response is given on that contra-factual scenario, psychological pressure is placed on the respondent not to give false information on actual incomes earned. In other words the idea is that a declaration on factual income as a head of a micro business would tend to be lower in the absence of a question first on the cost of opportunity in remaining independent.

Information on total sales and thus gross incomes are obtained from the respondent as well as current expenditures and salaries paid during the same period. This information is used to reconstruct by subtraction the mixed income of the individual leading the activity. This is Approach 2. Of course it is possible to obtain negative values either because that actually

[13]As an incentive to enlist respondent patience in the first phase, interviewers gave a gift of a detailed map of the country to respondents in those dwellings with eligible individuals and also a “statistical agenda”. In most cases both gifts were quite appreciated.

occurs during the period of reference (a calendar month used as reference) or simply because the information on total sales is underreported. This tends to be the Achilles Heel of the approach.

Before asking about total sales, it is important first to have an idea of the total current expenses a business has had during a year. To do this without relying on a single response, the questionnaire focuses on each type of expenditure and how many times a year a disbursement is made and the average monetary amount involved in each operation. This provides a very detailed picture of business expenditures which is a good basis for determining the actual level of activity. Among the economic items collected, expenditures are the first to be secured during the interview because the respondent has less incentive to underreport these amounts than the amount of total sales or income. Then the questionnaire focuses on the three main products or services provided, their unitary prices and quantities traded as well as the specific costs in trading them. From this information a margin or coefficient can be applied to the annual estimates of total current expenditures in case there is no information on total sales, as demanded in Approach 2. Hence Approach 3 consists of generalizing coefficients or margins from the main products to the whole activity. Subtracting wages from the total amount is another way to estimate the owner's mixed income, an approach which minimizes the risk of ending up with negatives figures.

Of all the variables in the survey it is most difficult to obtain accurate information on income of the micro-business. For each record in ENAMIN there are three possible ways to obtain this information. With data on the interest paid, it is possible to have for each production unit its operating surplus, and then with the addition of wages paid a reconstruction of the total value added (taking aside taxes and subsidies). If one approach is affected by a partial non-response, there is always at least another that can be used. If there is enough information to complete two or three approaches, the algorithmic rule of thumb in processing each record is to take the approach from where the maximum mixed income results, for it is assumed that the general tendency of respondents is to underreport their income.

These decisions are made at the end of the process. Before these calculations are made, all quantities are subject to double entry verifying to exclude outliers due to mistakes in digitalizing data. This also implies that consistency checks are made, most of them defined under deductive rules. However as described above, the hardest of all decisions are those dealing with total non- response which, as is shown in figure 6.9, is not negligible.

Changes made in the Mexican LFS (ENOE, see Chapter 4) provided additional information on those engaged in micro-businesses as well as contextual information and auxiliary variables to supplement information collected in ENAMIN. This provided new possibilities for different approaches to making imputations especially on quantitative variables of economic significance.

The imputation literature (Little & Rubin, 1986) classifies different techniques in two broad categories: deterministic and probabilistic. Among the former category, imputations of (i) general averages, (ii) averages by classes and (iii) conditional averages are the most common. In the case of the 2008 mixed modular wave, the procedure followed was to impute averages by building classes of affinities. A record in the first phase not interviewed in the second reflects the characteristics of the class of records in ENAMIN where it could belong. In principle this is a simple procedure however sophisticated by the criteria and methods used to define what similarity means.

One approach to the imputation procedure described above is to optimize the distance among records (observations) in a mathematical space as a function of the auxiliary variables selected to form clusters that the objective function guarantees are the most homogeneous among all possible combinations of observations. Statistical methods such as *"K Means"* or *Mahalanobis* distances are helpful in forming clusters defined in a multivariate way and optimized by means of an objective function.

There are other deterministic techniques such as those in which a "donor" record is selected and its value is taken and placed in the receptor record. If the donor is part of what already is contained in the survey, the technique belongs to the so called "hot deck" family; if it comes from another source it is called "cold deck"; both techniques stratify possible donors by alternative procedures.

An objection raised concerning deterministic methods is that if a data base is edited, then variances are going to be underestimated and the nature of the variable distribution will be modified. In particular imputation procedures by means of averages conspicuously alter multi-modal distributions. For this reason, probabilistic methods are increasingly suggested, either through introducing a random component in the selection of a donor register (Simple Random Hot Deck and Random Hot Deck by Classes) or through introducing a stochastic component in a regression. This last modality may help in maintaining the statistical precision needed to build intervals and test hypotheses, but of course it requires sound modeling which is not guaranteed as any experienced econometrician knows. For this reason, INEGI opted for imputation procedures which are not dependent on modeling and do not modify the micro-data offered to the public based on these imputations. The deterministic imputation described above and used in the various ENAMIN waves was only concerned with aggregates and is undertaken to supply national accountants with those economic aggregate variables that in turn become the starting point of their own procedures, as explained in Chapter 9.

3.4 The Survey of Employers and the Self-Employed in South Africa (SESE)

In South Africa the informal sector contributes around 6% to Gross Domestic Product. It also contributes 16.6% of the total employment. Following the international standard definition of the informal sector, these businesses are not registered with government. Without registration it is not possible to build a list of all such businesses (include them in a business register) and so they cannot be surveyed as part of a business or economic survey program.

As a result, the measurement of characteristics of informal sector businesses in South Africa is done every three years through a 1-2 survey methodology; the first stage involves identifying individuals who are running any kind of business through a household survey. In the case of South Africa this is the Quarterly Labour Force Survey (QLFS).14 The QLFS therefore provides the frame for the second stage of this

[14] The LFS (conducted twice per year) was replaced by the Quarterly Labour Force Survey in 2008. SESE was conducted twice using the LFS (March 2001, September 2005) and once (third quarter 2009) using the QLFS.

methodology, which is a follow up interview. This interview is done with individuals who are identified in the QLFS questionnaire as running businesses to determine the nature of their business and other characteristics of their businesses. The follow up interview is done through a survey called the Survey of Employers and the Self-Employed (SESE).

This paper therefore discusses the measurement of the informal sector in South Africa using the 1- 2 methodology, the definition of the informal sector, the challenges faced with the 1-2 survey methodology and lessons learnt in South Africa that may be useful for other countries intending to use the 1- 2 survey methodology.

3.4.1 The first stage of the 1 – 2 methodology

The first stage of the 1-2 methodology comprises a labour force survey that has a household sample that represents the whole country. The following questions in the labour force survey, which are used to establish employment, are also used to screen into SESE those who own businesses. (These questions are the very first of the labour market activity questions in the QLFS questionnaire.)

2.4	**In the last week….**	**YES NO**
	(a) Did you work for a wage, salary, commission or any payment in kind (including paid domestic work), even if it was for only one hour?	☐ 1 ☐ 2
	Examples: a regular job, contract, casual or piece work for pay, work in exchange for food or housing, paid domestic work.	
	(b) Did you run or do any kind of business, big or small, for yourself or with one or more partners, even if it was for only one hour?	
	Examples: Commercial farming, selling things, making things for sale, construction, repairing things, guarding cars, brewing beer, collecting wood or water for sale, hairdressing, crèche businesses, taxi or other transport business, having a legal or medical practice, performing in public, having a public phone shop, etc.	☐ 1 ☐ 2
	(c) Did you help without being paid in any kind of business run by your household, even if it was for only one hour?	
	Examples: Commercial farming, help to sell things, make things for sale or exchange, doing the accounts, cleaning up for the business, etc.	☐ 1 ☐ 2
	If yes to any part of Q 2.4 go to Section 4, otherwise go to Q 2.5	

To identify those running a business who were temporarily absent from work in the QLFS reference week, the following questions are used:

2.5	**In the last week, even though you did not do any work for pay, profit or did not help without pay in a household business, ….** **(a) Did you have a paid job that you would definitely return to?** ***→ If yes, go to Q 2.7, otherwise continue*** *Examples: a regular job, contract, casual or piece work for pay, work in exchange for food or housing, paid domestic work.* **(b) Did you have a business that you would definitely return to?** ***→ If yes, go to Q 2.7, otherwise continue*** *Examples: Commercial farming, selling things, making things for sale, construction, repairing things, guarding cars, brewing beer, collecting wood or water for sale, hairdressing, crèche businesses, taxi or other transport business, having a legal or medical practice, performing in public, having a public phone shop, etc.*	**YES NO** ☐ 1 ☐ 2 ☐ 1 ☐ 2

The advantage with this screening process is that it takes into consideration that a person could be in paid employment but also running a business, so the design allows for multiple activities to be filled in.

3.4.2 The second stage

A SESE questionnaire is administered to anyone identified as running a business in the QLFS questionnaire. The first question asked of business owners in SESE is, if they run any kind of business big or small for themselves or with one or more partners. This question is asked to business owners who are already identified in the QLFS because of the time lag between the QLFS and the SESE interview as some of the small businesses sometimes close down before the SESE interview can take place. The SESE interview is usually conducted two weeks after the LFS interview.

Statistics South Africa aligned its definition of the informal sector enterprises to the definition adopted at the 15th ICLS and uses registration for VAT and Income Tax as the main determinants in defining enterprises as being in the formal or the informal sector. In essence all household businesses that registered for either VAT or Income Tax in the SESE are regarded as formal sector businesses.

Accordingly, the next question that is asked in the SESE pertains to registration of the enterprise unit. Business owners are asked whether their business is registered for VAT or Income Tax. This question is asked because at the identification stage in the QLFS (questions 2.4 and 2.5 above) no distinction is made between formal and informal sector businesses. However, it must be mentioned that in the QLFS a question on registration for VAT and Income Tax is asked of employers and own account workers. The question is not used for screening purposes for SESE because in the QLFS only details of the main

activity are recorded and it is possible that some individuals may be running a business and engaged in wage employment. There is opportunity however for countries that collect details of all activities individuals are engaged in, to screen for registration at the LFS stage.

Once a business is reported as registered for VAT then it is the end of the SESE interview for the business owner. However, businesses that are registered only for income tax are, together with household enterprises that are not registered, taken through to questions on SESE. Leaving in income tax registered businesses provides time-series continuity as the previous SESEs used only VAT registration in the definition of informal sector enterprises. However, during the analysis stage, all the enterprises registered for income tax will be analysed as not being in the informal sector.

The questions that are asked in the SESE are enterprise-specific and they relate to goods and services provided by the business, ownership of business, source of funding for starting the business, location, record of accounts, size of business (employees), reasons for starting the business, duration of operation, operational costs, turnover and labour costs. A copy of the SESE questionnaire is included as Annex 3. Since individuals may run more than one business in parallel, the questionnaire allows for the recording of information for up to three businesses.

3.4.3 Estimation process

If all individuals identified as target population15 for the follow-up survey respond during the follow-up interview, then the weights in the first interviews will be sufficient because the reporting is based on individuals in the target population. However, not all identified individuals respond during the follow-up for various reasons. Adjustment for non-response could be done using one of the following approaches:

1. The computed weights for all qualifying persons from the QLFS are adjusted for person non-response in SESE to get the final SESE weights. Basically, the product of the adjustment factors and LFS person weights of the selected records yield the SESE weights.
2. The second approach is to impute the whole record of the non-responding individual using similar-record substitution and keep the weights from the first interview.

Approach 1

This is the approach that South Africa was using for previous SESEs which were conducted as a supplement to the LFS. The general file that contains information about the response codes for the follow-up interviews was linked to the Labour Force Survey. This was to make sure that the main demographic variables of interest such as gender, age, and population group were kept for benchmarking process. The response rate was calculated by using the response codes from the follow-up questionnaires.

[15] Running a business which is not registered for VAT in the case of South Africa.

Individual records were classified in the three categories 'response', 'non-response' and 'out-of-scope'. The out-of-scope category included individuals who reported that they were in the target population during the LFS interview, but on verification during the follow-up they were no longer running a business.

The following adjustment classes were created: Province, District councils and demographic variables (age group, population group and gender).

In general, the response rate was given by the weighted response divided by the weighted sum of response and non-response in the follow-up interview within an adjustment class. The non-response adjustment factor is the inverse of the response rate. The out-of-scope units in the follow-up interviews were ignored during the calculation of the adjustment factors. The adjustment factors were applied to the LFS person weights in order to obtain the final weight for the follow-up interview on SESE.

Approach 2

This approach takes the advantage that a great deal is known about the non-respondents, specifically, all of the information that they have provided to the labour force survey. There is no need to depend on the assumption that geographic proximity implies personal similarity as it is the case in Approach 1. Instead, to compensate for non-response similar record substitution can be used. It is important to establish what is "similar", by using as much of the QLFS questionnaire as is mathematically feasible. (The more variables used to define "similarity" the greater the likelihood that an exact match may not be found and matching criteria would have to be relaxed.)

Statistics South Africa revamped its processing systems making it possible to adopt this approach for estimation of future SESEs which are conducted in conjunction with the South African QLFS.

It is important to note that the system requirements for such a non-response adjustment algorithm are substantial and not all national statistical offices will have the necessary infrastructure.

3.4.4 Challenges

One of the challenges is the risk of double-counting when business owners, who are in partnership, regard themselves as equal partners but do not reside in the same household. The probability of double-counting increases with the geographic proximity of the partners. If they live in the same primary sampling unit, the probability of double-counting is much higher than if they live in different primary sampling units. Given that SESE addresses the informal sector, proximity of the partners is quite likely. However, in South Africa there is only 1.9% of those who run unregistered businesses who fall into this category.

The longer the time lag between the first stage and the second stage the greater the chance of finding out that some businesses, that were identified in stage one, have ceased to exist while others may have moved out of the sampled households. However, this may be mitigated by informal sector enterprises that come into existence during the time lag, provided those are established in the sampled households. In the case of South Africa the

time lag between the two stages is two weeks and therefore this scenario is not much of a challenge. The only way to avoid the time lag effect is to keep stage 1 and stage 2 as close a possible during the enumeration process. However, one has to be careful with the resulting response burden.

It is also possible that sometimes business owners may not be able to separate their business expenses from their household expenditure, which may lead to an overestimation of their business expenses and an underestimation of their value added and operating surplus. For each type of expenditure items, the SESE questionnaire asks to indicate if the expenses for the business can be separated from household expenditure. If this is not the case, respondents are to report the full amount (business and household). Business owners may also sometimes not know what their turnover is, especially if the reference period is twelve months. In the case of South Africa the reference period for reporting on expenses and turnover is the last month. An assumption is made that the month in which this information is collected represents all the months of the year and thereby annual estimates are computed on the basis of this.

In countries, where the informal sector is big, the requirements for national accounts to obtain data per each economic activity branch can probably be met adequately using the 1-2 methodology as the sample from the first stage (LFS) will be big enough to yield good estimates for the second stage (SESE). However, because in the case of South Africa the informal sector is relatively small the sample from the first stage is not big enough to disaggregate to industry level for certain variables such as turnover. For example in the 2005 SESE the co-efficient of variation (CV) on turnover for different industries ranged between 14% and 51%. However the CVs for turnover at the aggregated level were reliable at around 14%. In South Africa turnover at industry level is not published because of the low levels of precision.

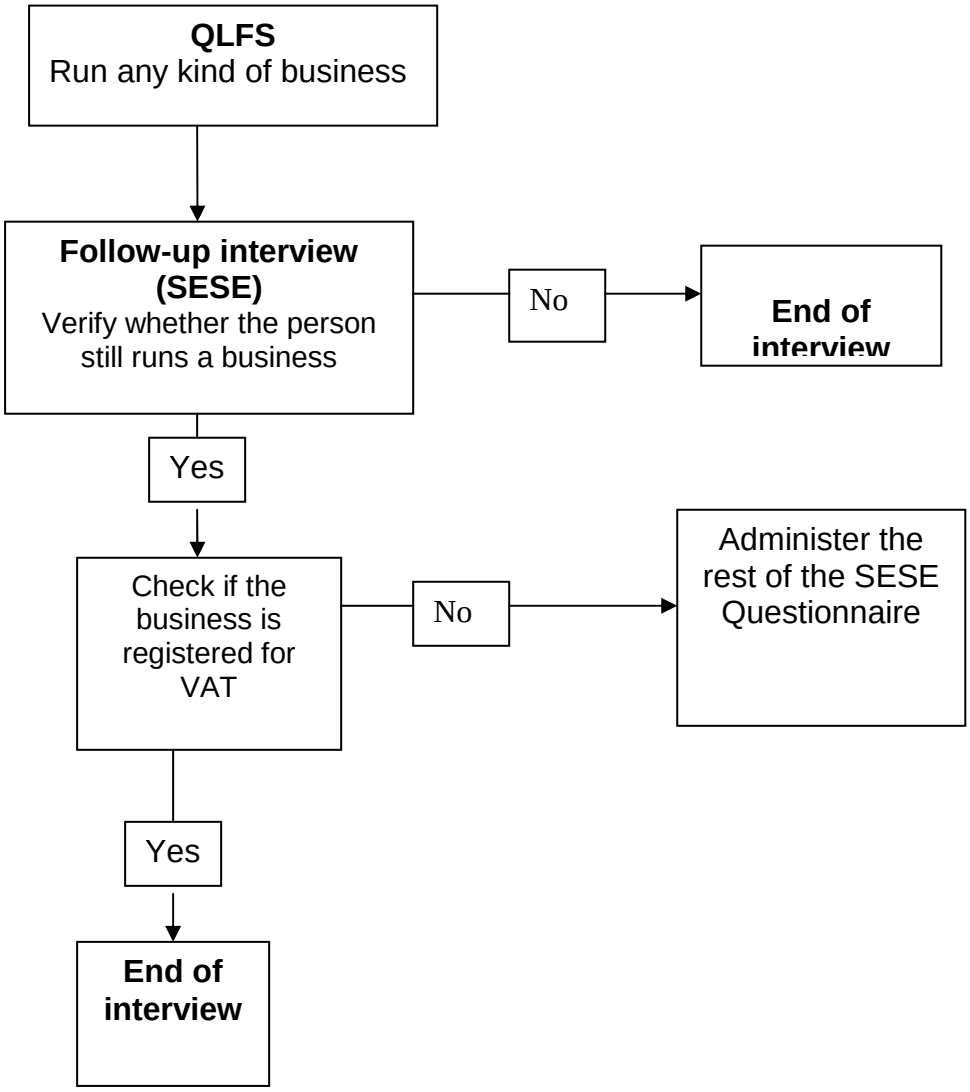
It is important that as far as possible the interviews are carried out with the enterprise owners so as to minimize the rate of missing information and misreporting. In the case of South Africa the fieldworkers are trained not to interview proxy respondents but to always interview the business owner. Any information that is missing is due to the owner not knowing how, or not wanting, to provide the information and is not due to any proxy effect.

3.4.5 Conclusion

Unregistered businesses have zero probability of being in the samples of any of the business surveys; it is because of this reason that Statistics South Africa uses the 1-2 methodology to collect information about the unregistered businesses. One of the main defining characteristics of these businesses that make them suitable for this 1-2 approach is that they are numerous which improves the efficiency of any sample design.

The link between the labour force survey and the SESE makes it possible to profile business owners in terms of other information collected in the QLFS such as their demographic characteristics and key labour market indicators. The 1-2 methodology in South Africa has worked quite well as there was no need to list businesses separately from the normal household list.

Diagram of the 1-2 methodology



4. The LSMS experience in measuring the informal sector

Even if it is not their main objective, the Living Standard Measurement Studies (LSMS) surveys can be considered as a form of modular mixed informal sector survey. Launched by the World Bank in the mid 1980s, the LSMS programme was initially designed to measure and monitor poverty and inequality, but interest in modelling the micro-economic behaviour of households has led to changes in their structure. Based on relatively small sample sizes at national scale (e.g., 1,600 households in Côte d'Ivoire), they have become multi-purpose studies, covering almost all aspects of the economic and domestic activities of households: consumption, income, agriculture, jobs, transfers, migration, education, health, anthropometry, etc. In terms of surveys, this programme registered a new impetus in the 1990s to meet information needs for two activities in which the World Bank has been deeply involved: the reconstruction of statistical systems in transition countries and the launching of the new international poverty reduction strategies.

In most of the LSMS survey questionnaires, the informal sector and informal employment can be tracked through the "*Employment*" and "*Non-Farm Household Business*" (*NFHB*) modules. The "*Employment*" module is conceived as a simplified LFS, where core labour market indicators are collected. For wage workers (only), detailed information on wage earnings is collected as an element of the characteristics of jobs, which can be used to measure the wage component of informal employment. Information on *NFHB* is collected in the corresponding module (with the same procedure as for Farm Household Business). Each surveyed household is asked if it possesses at least one non-farm business. If yes, the screened households are supposed to answer some questions about the *NFHB's* characteristics, mainly to provide information on income drawn from these activities. In some questionnaires, up to four *NFHBs* per household are considered. A distinction between formal and informal *NFHBs* can be made if the module includes questions on the business size or registration (which is not systematically the case). From an informal sector survey perspective, the LSMS survey is a modular mixed survey, where phases 1 and 2 are integrated in the same questionnaire and household production units are selected with a probability equal to 1.

Nevertheless, for informal sector measurement purposes the LSMS surveys present various shortcomings. Basically, these shortcomings result from the LSMS surveys not being designed for measuring the informal sector, which is only a by-product of the survey. Apart from small sample sizes, two main weak points can be noted: a) *limited reliability*: questions related to production and income are not formulated in as a detailed manner as it should be to capture the informal sector aggregates; b) *partial subject specific coverage*: some important indicators are not considered in the questionnaire (origin of inputs, destination of production, investment, capital, prices, difficulties and demands, etc.). Furthermore, the link between jobs ("*Employment*" module corresponding to phase 1) and production units ("*NFHB"* module corresponding to phase 2) is not systematically established. Moreover, it is not always possible to identify the informal status of household businesses (for instance, the question on registration was not considered in the first LSMS surveys). If we add the fact that the labour market indicators are not necessarily collected in line with international standard (ILO) definitions, and the great complexity of deriving (approximated) informal sector indicators,[16]

[16] An exercise conducted in Vietnam shows that it is necessary to manipulate 10 different files from the Vietnamese Household Living Standard Survey to compute an indicator of informal sector employment, while

the LSMS surveys can only be regarded as a second-best strategy to measure informal sector and employment.

In spite of their shortcomings, the LSMS surveys present some advantages for our objectives. First, in many countries where no specific informal sector survey has been conducted, the LSMS survey is probably the best available source on informal sector, informal employment and related issues. Second, given their extensive module coverage and multipurpose character, the LSMS surveys can be used to explore the link between the informal sector and other issues (such as health, education, migration, poverty etc.). Finally, an LSMS survey could be used as the filter survey (phase 1) to select the sample for phase 2 (and eventually phase 3) of a specific modular mixed informal sector survey. In such a case, the questionnaire should be revised and completed. However, the risk of such an option is to overload an already heavy survey by adding new modules. Moreover, the size of the filter survey sample may have to be increased and its design modified.

5. Outlook

This chapter has shown that the modular mixed approach is a robust and proven method to meet the requirement of data collection on the informal sector and informal employment, albeit in need of further refinement. One may even state that it is nowadays easier in some countries to obtain reliable estimates for the informal sector than for the formal sector. The basic reasons are twofold. On the one hand, the modular mixed surveys have overcome the main coverage error shortcomings of the traditional surveys on the informal sector. On the other hand, measurement errors are usually lower than in formal sector surveys, due to the higher willingness of the former to answer the survey questions. Three decades of accumulated experience in the field provide enough empirical evidence for considering the modular mixed survey as an instrument to be included in the standard toolbox of survey statisticians.

The main task facing survey statisticians today is not to devise new procedures, but to disseminate experiences that have proved their effectiveness so that they can be put into general use. Among the further issues to deal with, two major ones can be pointed out, which have both technical and institutional implications but address the same key point: how to build up a sustainable survey system to monitor informal sector and informal employment over time?

From a statistical perspective such a system cannot be based on one-shot surveys (as it has often been the case in the past) but on a set of surveys comparable in the long run. The ideal system should rely on four types of surveys:

- a (revised) LFS to address labour market and informal (sector) employment issues;
- an Informal Sector Survey (or Household Enterprises Survey) to address supply-side issues (production, etc.);
- a (revised) HIES to address demand-side issues (household consumption, etc.);
- a (revised) system of price indices to address the issue of price evolution in the informal sector.

---

this number increases up to 14 files for the calculation of income; strong assumptions have to be made to assign jobs to the informal sector (Razafindrakoto and Roubaud, 2007).

In general, the current LFS should systematically address informal sector and employment issues (both to provide indicators and, possibly, to serve as filter survey for phases 2 and 3). One important implication of this strategy is to adjust the sampling design to this purpose. Too often, the LFS sample is designed to optimize the estimation of the unemployment rate. This makes (reasonably) sense in developed countries but is not adapted to the situation in developing countries. All accumulated evidences (theoretical and empirical) suggest that unemployment is not a sufficient indicator to capture labour market tensions in LDCs, but informal sector and employment indicators are. It is time to take this fact into account and adapt the LFS sample designs accordingly (the two objectives are not necessarily competing).

Specific price indices should also be elaborated. It is the only way to deflate the informal sector aggregates at constant prices, as no other price indices currently collected (CPI, PPI) can be considered as a good proxy of informal sector prices. In market economies, prices are the basic mechanism for competition and resource allocation. In analytic terms, reliable information on informal sector prices is essential to understand price formation and dynamics. Using the information collected in phases 2 and 3 for weighting, specific price indices can be designed and prices collection could be organized adjusting already existing operations for the CPI or the PPI. The *1-2-3 survey* provides a reasonable alternative (mobilizing unit prices/values captured in phases 2 and 3). It fits also with the global objective of the set of four surveys mentioned before, unless data requirements or statistical infrastructures call for stand-alone surveys.

The institutional channels need to be organized through which modular mixed surveys on the informal sector can be permanently integrated into the national economic information systems. Upstream, there is a need for close cooperation between the survey statisticians responsible for collecting data on the informal sector and the potential users, such as national accountants, starting with survey design. Downstream, the results of the surveys should be more broadly disseminated to different parts of the development community: national accounts services have already been mentioned, but policy makers, academia and researchers are also strategic users to be considered.

References

J. Abowd and F. Kramarz (1999), "The Analysis of Labor Markets using Matched Employer-Employee Data", in O. Ashenfelter and D. Card eds., *Handbook of Labor Economics*, chapter 26, volume 3B, North-Holland, 2629-2710.

Amegashie F., Brilleau A., Coulibaly S., Koriko O., Ouédraogo E., Roubaud F. and C. Torelli (2005), "La conception et la mise en œuvre des *enquêtes 1-2-3* en UEMOA, les enseignements méthodologiques", *Statéco* No.99, 21-41.

Brilleau A., Ouedraogo E. and F. Roubaud (2005), *L'enquête 1-2-3 dans les pays de l'UEMOA: la consolidation d'une méthode*, Revue Stateco No.99, AFRISTAT-DIAL-INSEE, 15-170.

Brilleau A. and F. Roubaud F. (2005), "Propositions pour la mise en place d'un dispositif de suivi de l'emploi et du secteur informel dans les Etats membres de l'UEMOA", *Statéco* No.99, 155-170

Cling J.-P., Razafindrakoto M. and F. Roubaud, eds. (2003), *New International Poverty Reduction Strategies*, London/ New York: Routledge.

DANE (1999), *Construcción de marco de áreas para las investigaciones económicas*, Bogota: DANE.

De Vreyer P. and F. Roubaud, eds. (2008), *Urban labour Market in Africa*, forthcoming.

De Soto, Hernando (1987), *El Otro Sendero*/ Diana, first edition, México

DIAL (2006), *The 1-2-3 Toolkit: didactic software and documentation library*, available on http:/www.dial.prd.fr.

ESCAP (2007), "A Unified Data Collection Strategy for Measuring: the Informal Sector and Informal Employment", Statistics Division, ESCAP, Bangkok, Thailand.

ESCAP (2006), "Project 0607A: Interregional Cooperation on the Measurement of Informal Sector and Informal Employment. Project Document", Statistics Division, ESCAP, Bangkok, August.

Herrera J., Razafindrakoto M. and F. Roubaud (2007), "Governance, Democracy and Poverty Reduction: Lessons drawn from household surveys in sub-Saharan Africa and Latin America", *International Statistical Review*, 75(1), 70-95.

Hussmanns R., Mehran F. and V. Verma (1990), *Surveys of Economically Active Population, Employment, Unemployment and Underemployment: An ILO Manual on Concepts and Methods*, Geneva: International Labour Office.

INEGI/ STPS, Encuesta Nacional de Micronegocios, 1992, 1994, 1996, 1998, 2002,

2008, México: www.inegi.org.mx

INEGI (2005), *Encuesta Nacional de Ocupación y Empleo: una nueva encuesta para México*,

www.inegi.org.mx

Levy Santiago (2008), *"Good intentions, Bad Outcomes: Social Policy, Informality and Economic Growth in Mexico"*, Brookings Institution Press, Washington

Little Roderick J.A. and Rubin D.B., *Statistic Analysis with Missing Data*, John Wiley and Sons, New York

Maloney William F., (1998), *Are labor markets in developing countries dualistic?*, Policiy Research Working Papers, WPS 1941, World Bank, Washington

Mayorga Mora E.A. and I.I. Parra Ramírez (2004), "Diseño metodológico", in Herrera J. et al. op.cit., , 67-87.

Ordaz E. and F. Roubaud (1989), "*Documento metodológico de la investigación sobre el sector informal en México*", communication presented at the COM-SIES Conference, Santiago, Chile.

Organisation of American States (1992), "*Seminario interamericano sobre medición del sector informal*", Lima, Peru.

Rakotomanana F., Ramilison E. and F. Roubaud (2003), "The creation of an annual employment survey in Madagascar. An example for sub-saharan Africa", *InterStat* No.27, September, 35-58.

Rakotomanana F., Ravelosoa R. and F. Roubaud (2003), "The 1-2-3 survey of the informal sector and the satisfaction of household needs in the Antananarivo conurbation", *InterStat* No.27, September, 59-88.

Razafindrakoto M. and F. Roubaud (2008), "The Informal Sector in Viet Nam: What do we know and How to go further? A statistical and economic perspective", *Thong tin* No.2008/1-2, 110-135.

Razafindrakoto M. and F. Roubaud (2003), "The existing systems for monitoring poverty: weaknesses of the usual household surveys", in Cling J.-P. et al., op cit, Chapter XI, 265-294.

Roubaud F. (1994), *L'économie informelle au Mexique : de la sphère domestique à la dynamique macro-économique*, Paris: Karthala/Orstom.
(translation into spanish, *La economía informal en México : de la esfera doméstica a la dinámica macroeconómica*, México: Fondo de Cultura Económica).

Roubaud F. and Séruzier M. (1991), *Economie non enregistrée par la statistique et secteur informel dans les pays en développement*, Stateco No.68, INSEE, Paris.

Vescovo A. (2007), "*La mesure des revenus et du niveau de vie des ménages: cohérence interne des enquêtes 1-2-3 de sept capitales économiques de l'UEMOA "*, Bamako, Mali: AFRISTAT.

Villamizar M.E. (2004), "*Algunos antecedentes de la medición del sector informal*", in Herrera J. et al. op.cit., Chapter 2, 63-66.

ANNEX 1

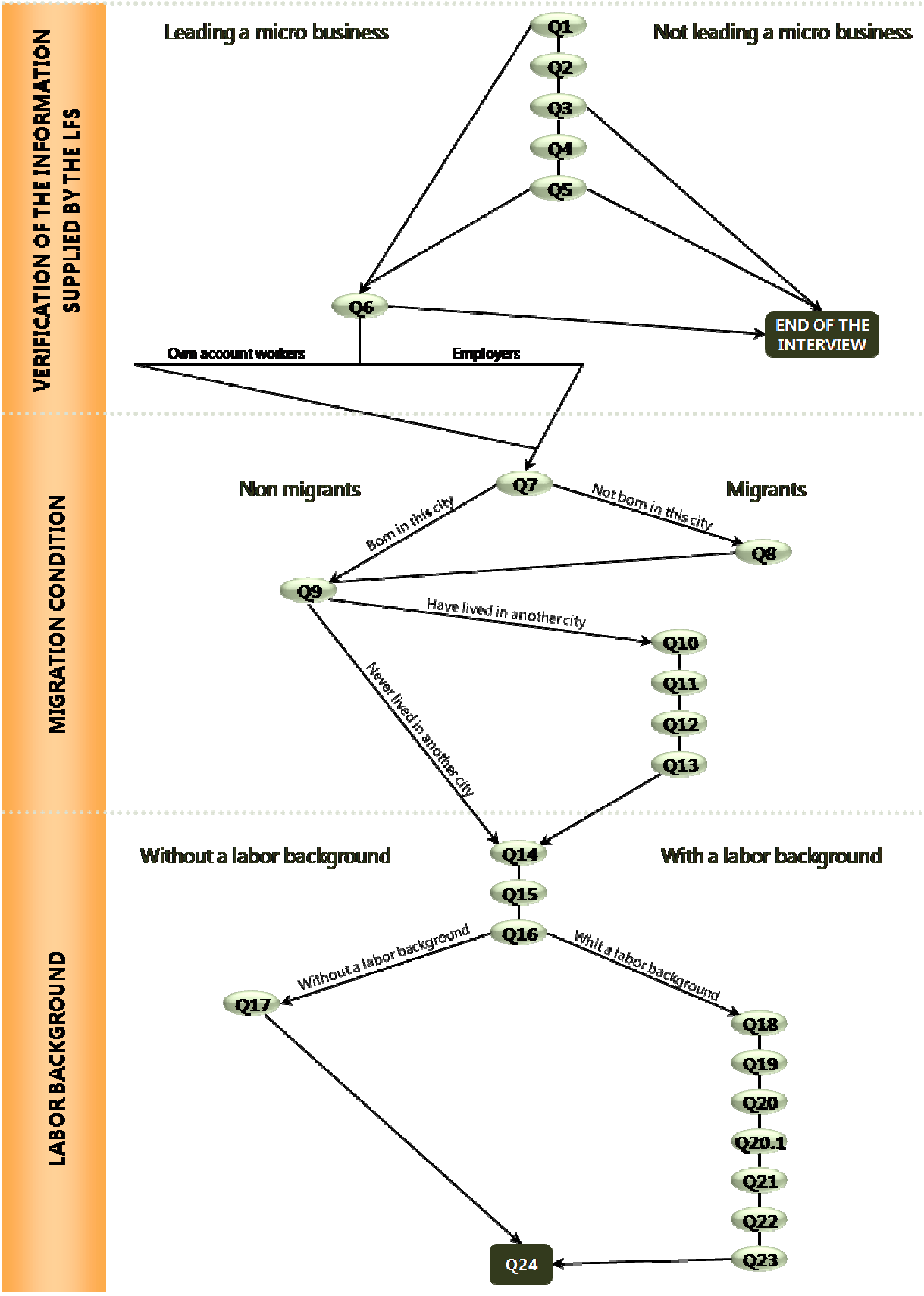
UN Development Account Project 'Interregional Cooperation on the Measurement of Informal Sector and Informal Employment'

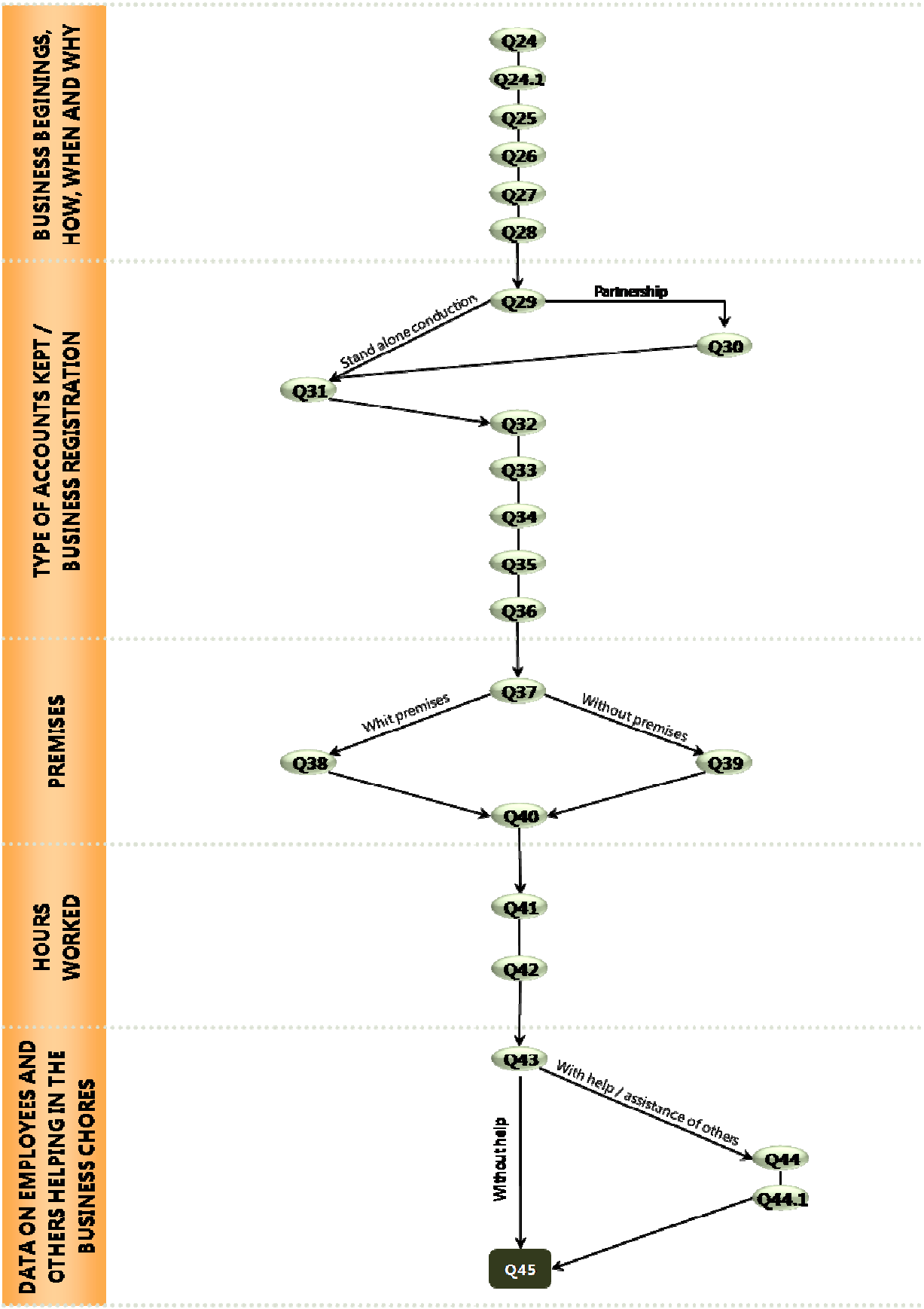
SURVEY QUESTIONNAIRES

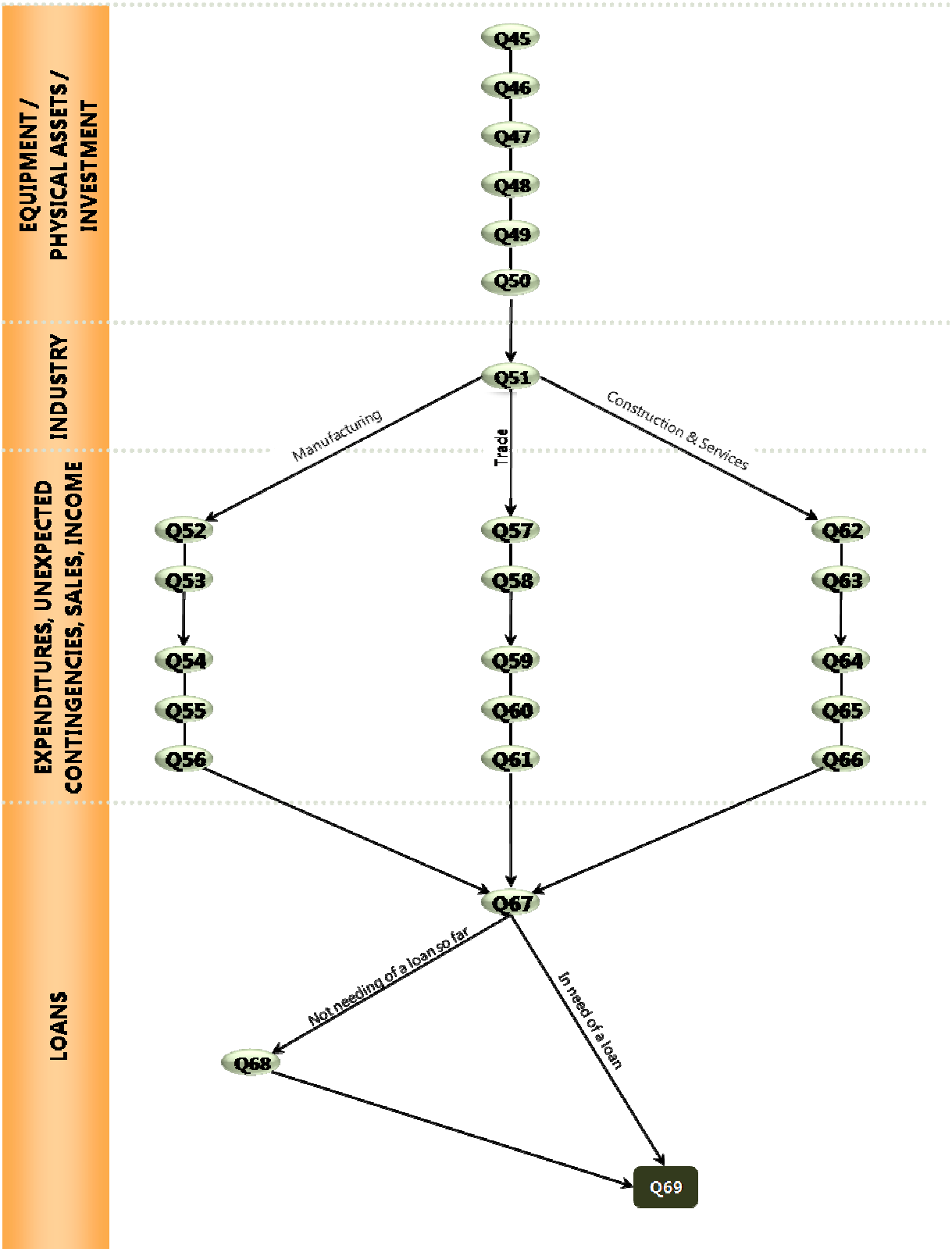
A. Phase 1
B. Phase 2

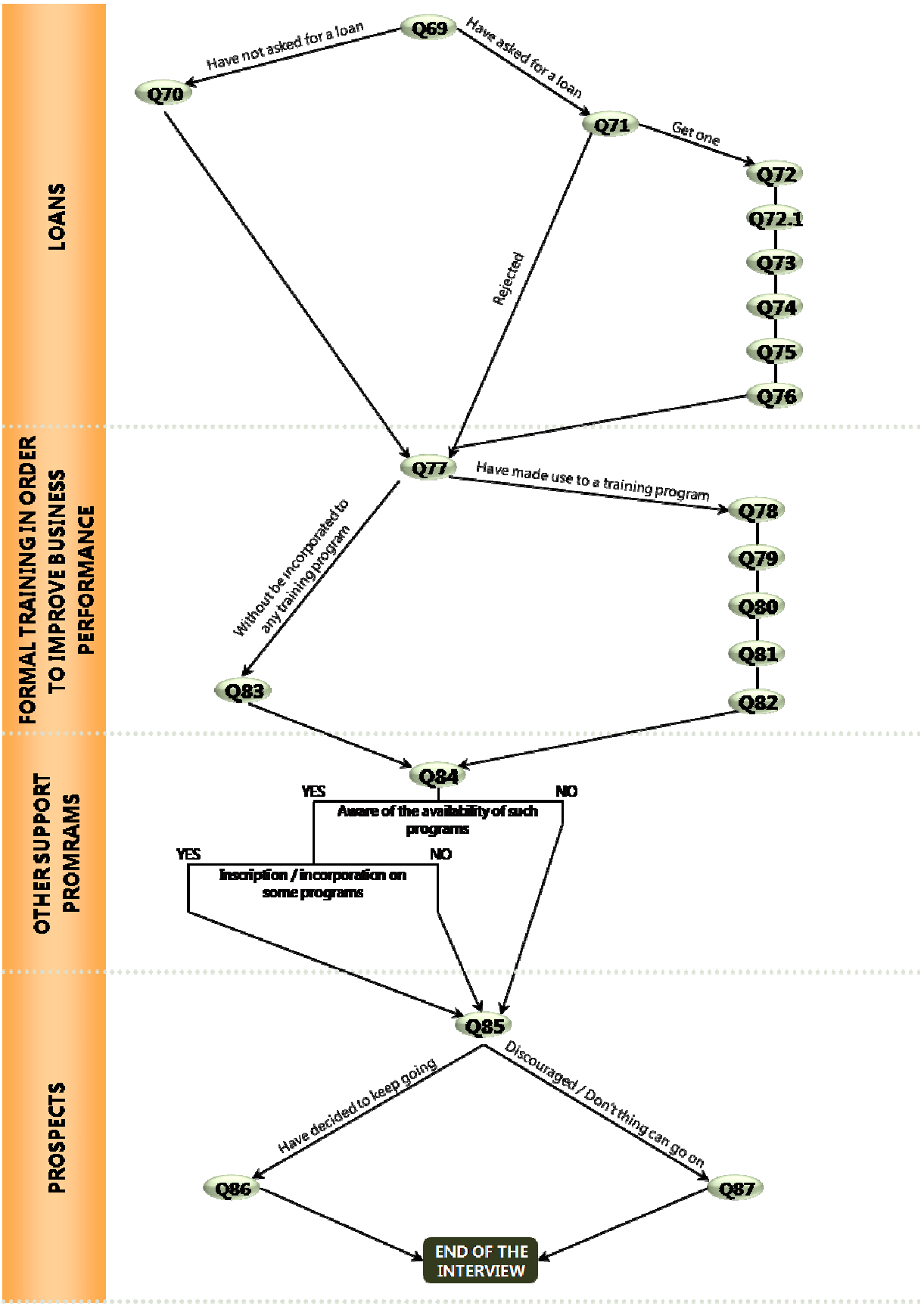
ANNEX 2

STRUCTURE OF THE ENAMIN QUESTIONNAIRE


VERIFICATION OF THE INFORMATION SUPPLIED BY THE LFS
Leading a micro business
Not leading a micro business
Q1
Q2
Q3
Q4
Q5
Q6
END OF THE INTERVIEW
Own account workers
Employers
MIGRATION CONDITION
Non migrants
Q7
Migrants
Born in this city
Not born in this city
Q8
Q9
Have lived in another city
Never lived in another city
Q10
Q11
Q12
Q13
LABOR BACKGROUND
Without a labor background
Q14
With a labor background
Q15
Q16
Without a labor background
Whit a labor background
Q17
Q18
Q19
Q20
Q20.1
Q21
Q22
Q23
Q24


BUSINESS BEGINNINGS, HOW, WHEN AND WHY
TYPE OF ACCOUNTS KEPT / BUSINESS REGISTRATION
PREMISES
HOURS WORKED
DATA ON EMPLOYEES AND OTHERS HELPING IN THE BUSINESS CHORES
Q24
Q24.1
Q25
Q26
Q27
Q28
Q29
Partnership
Stand alone conduction
Q30
Q31
Q32
Q33
Q34
Q35
Q36
Q37
Whit premises
Without premises
Q38
Q39
Q40
Q41
Q42
Q43
With help / assistance of others
Without help
Q44
Q44.1
Q45


EQUIPMENT / PHYSICAL ASSETS / INVESTMENT
INDUSTRY
EXPENDITURES, UNEXPECTED CONTINGENCIES, SALES, INCOME
LOANS
Q45
Q46
Q47
Q48
Q49
Q50
Q51
Manufacturing
Trade
Construction & Services
Q52
Q53
Q54
Q55
Q56
Q57
Q58
Q59
Q60
Q61
Q62
Q63
Q64
Q65
Q66
Q67
Not needing of a loan so far
In need of a loan
Q68
Q69


LOANS
FORMAL TRAINING IN ORDER TO IMPROVE BUSINESS PERFORMANCE
OTHER SUPPORT PROMRAMS
PROSPECTS
Q69
Have not asked for a loan
Have asked for a loan
Q70
Q71
Get one
Q72
Q72.1
Q73
Q74
Q75
Q76
Rejected
Q77
Have made use to a training program
Q78
Q79
Q80
Q81
Q82
Without be incorporated to any training program
Q83
Q84
YES
NO
Aware of the availability of such programs
YES
NO
Inscription / incorporation on some programs
Q85
Have decided to keep going
Discouraged / Don't thing can go on
Q86
Q87
END OF THE INTERVIEW

ANNEX 3

SESE 2009 QUESTIONNAIRE

Statistics South Africa
Preferred supplier of quality statistics

Survey of Employers & Self-Employed 2009

A. Particulars of dwelling

A1. PSU number

A2. Dwelling unit number

A3. Assignment number

A4. Survey date

A5. Physical identification of the dwelling unit

A6. Telephone no. of enumerated household

A7. Household number for this household

Q8. Questionnaire number

A9. Total number of questionnaires for this household

A10. Person number (column no. from QLFS questionnaire)

A11. Respondent name

B. Field staff

B1. SO		Persal no.		Int. date	
B2. DSC		Persal no.		Date chkd	
B3. PQM		Persal no.		Date chkd	

Unique no.

C. Response details

Visit no	Date (actual) d d m m y y y y	Result code	Next visit (planned) d d m m y y y y
C1.			
C2.			
C3.			
C4.			
C5. FINAL RESULT			

C6. Comments and full details for result code 02 – 11

RESULT CODES

1	Completed
2	Non-contact
3	Refused
4	Partly completed
5	No usable information
6	Vacant /unoccupied dwelling
7	Listing error
8	Demolished
9	Change of status
10	Other non response
11	Ended at question 1

Comment in C6 giving full details for result code 02 - 11

	h	h	m	m
INTERVIEW START TIME				

INTRODUCTION

Read out: The last time we spoke to you or a member of your household, we found out that you were engaged in some business activities. Statistics South Africa has a great deal of information about medium and large-sized businesses but very little information about small businesses like yours. More information about such businesses is needed for better government planning. All information you provide will be held strictly confidential and will not be made available to anyone else inside or outside of government.

1	**Do you run any kind of business, big or small, for yourself or with one or more partners?** **1 = YES** 2 = NO → ***End of interview***	☐ 1 ☐ 2
2	**Do you run more than one business?** **1 = YES** 2 = NO → ***Go to Q4***	☐ 1 ☐ 2
3	**How many businesses do you run?**	☐☐

If more than one business, ask which business has normally the highest turnover. Record that business as Business 1. Record the business with the second highest turnover as Business 2, etc. If only one business, complete the column for Business 1.

		Business 1	**Business 2**	**Business 3**
4	**Is your business registered for VAT?** ***(repeat for each business)*** 1 = YES 2 = NO	☐ 1 ☐ 2	☐ 1 ☐ 2	☐ 1 ☐ 2
5	**Is your business registered for income tax?** ***(repeat for each business)*** 1 = YES 2 = NO	☐ 1 ☐ 2	☐ 1 ☐ 2	☐ 1 ☐ 2
6	**Do you have any licenses or permits to operate this business?** 1 = YES 2 = NO	☐ 1 ☐ 2	☐ 1 ☐ 2	☐ 1 ☐ 2

		Business 1	Business 2	Business 3
7	**Who issued the license/s or permit/s?** 1 = Municipality/ Provincial authority ……..…….. 2 = Professional association …….. 3 = Business association ……..…….. 4 = Regional Services Council ……..…….. 5 = Traditional leader ……..…….. 6 = Protection agency/ies ……..…….. 7 = Friend/relative ……..…….. 8 = Other, *specify*	YES NO 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2	YES NO 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2	YES NO 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2
8	**Is your business registered for Unemployment Insurance Fund, UIF?** 1 = Yes 2 = No	1 2	1 2	1 2
9	**In your business are you……** 1 = Own-account worker (without employees)? 2 = Employer (with employees)?	1 2	1 2	1 2
10	**What kinds of goods or services does the business provide?** *Describe in at least three words. If more than one activity, answer for the one which usually generates the most net income.*			
11	**What is the name of your business?** *Write 'No name', if relevant.* ***For coding***			

		Business 1	Business 2	Business 3
12	**Is the business owned by you as a single owner?** 1 = YES 2 = NO	☐ 1 ☐ 2	☐ 1 ☐ 2	☐ 1 ☐ 2
13	**Is the business owned in partnership with other members of your household?** 1 = YES 2 = NO	☐ 1 ☐ 2	☐ 1 ☐ 2	☐ 1 ☐ 2
14	**How many household members, including you, are partners in the business?**			
15	**Are there any business partners living in other households?** 1 = YES 2 = NO	☐ 1 ☐ 2	☐ 1 ☐ 2	☐ 1 ☐ 2
16	**How many of those partners live in other households?**			
17	**Who is the main owner of this business?** 1 = Yourself 2 = Another family or household member 3 = Another person in the partnership or cooperative, not a household member 4 = The ownership is equally shared between two or more owners 5 = Other, *specify*	☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5	☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5	☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

Questions about the site for operation

		Business 1	Business 2	Business 3
18	**Does this business operate ….** 01 = Within the owner's dwelling/s – with its own space (e.g. a separate room) 02 = Within the owner's dwelling/s – without its own space (e.g. a family room) 03 = In a structure attached to owner's dwelling/s or on the same plot (e.g. a workshop in the back yard) 04 = Within another person's dwelling (e.g. a neighbour's dwelling) 05 = In a non-residential building (e.g. an office block or factory) 06 = From a taxi rank / bus station / train station 07 = On a footpath, street or open space 08 = At a market 09 = No fixed location/mobile 10 = At customer's homes or offices → ***Go to Q22*** 11 = Other, *specify*	☐ 01 ☐ 02 ☐ 03 ☐ 04 ☐ 05 ☐ 06 ☐ 07 ☐ 08 ☐ 09 ☐ 10 ☐ 11	☐ 01 ☐ 02 ☐ 03 ☐ 04 ☐ 05 ☐ 06 ☐ 07 ☐ 08 ☐ 09 ☐ 10 ☐ 11	☐ 01 ☐ 02 ☐ 03 ☐ 04 ☐ 05 ☐ 06 ☐ 07 ☐ 08 ☐ 09 ☐ 10 ☐ 11
19	**Is the location of the business a permanent** (e.g. over a period of time) **or a temporary arrangement?** 1 = PERMANENT 2 = TEMPORARY 3 = NOT APPLICABLE	☐ 1 ☐ 2 ☐ 3	☐ 1 ☐ 2 ☐ 3	☐ 1 ☐ 2 ☐ 3
20	**Do you pay for use of this location for business purposes?** 1 = YES 2 = NO 3 = NOT APPLICABLE	☐ 1 ☐ 2 ☐ 3	☐ 1 ☐ 2 ☐ 3	☐ 1 ☐ 2 ☐ 3
21	**How much did you pay in the last calendar month? (Rands)**	☐☐☐☐☐☐	☐☐☐☐☐☐	☐☐☐☐☐☐

22	**What records do you keep for this business?** 1 = Simple informal records of sales and/or expenditures 2 = Some accounts but not full (for example expenditures) 3 = Full annual accounts 4 = No accounts kept	1 2 3 4	1 2 3 4	1 2 3 4

		Business 1	Business 2	Business 3
23	**Are the expenditures for the business recorded separately from expenditures for the household?** 1 = YES, THEY ARE ALL RECORDED SEPARATELY 2 = SOME ARE RECORDED SEPARATELY, SOME TOGETHER 3 = NO, THEY ARE ALL RECORDED TOGETHER 4 = NO, BUSINESS EXPENDITURES ARE NOT RECORDED	1 2 3 4	1 2 3 4	1 2 3 4
24	**How many months in the last twelve months did the business operate?** ***If <u>12 months</u> go to Q25*** ***If <u>less than 12 months</u> go to Q26***			
25	**In the last 12 months, how much money came into the business through sales or services offered, before any deductions (turnover)? (Rands)** → ***Go to Q27***			

26	Ask if less than 12 months in Q24, otherwise go to Q27 **What was the main reason that the business had no activity in some months?** 01 = SEASONAL FACTORS 02 = FAMILY REASONS (E.G. SICK CHILD) 03 = NON PAYMENT OF GOVERNMENT CHARGES/TAXES 04 = NO CUSTOMERS 05 = SICKNESS OF YOURSELF OR STAFF 06 = OTHER PERSONAL REASONS THAN SICKNESS (PREGNANCY, ETC) 07 = LACK OF RAW MATERIALS 08 = LACK OF FUNDS TO BUY SUPPLIES 09 = NO ONE TO HELP DURING OWNER'S ABSENCE 10 = BUSINESS CREATED DURING THE PAST 12 MONTHS 11 = DUE TO VIOLENCE OR CRIMINAL INCIDENT/S 12 = OTHER, *specify*	☐ 01 ☐ 02 ☐ 03 ☐ 04 ☐ 05 ☐ 06 ☐ 07 ☐ 08 ☐ 09 ☐ 10 ☐ 11 ☐ 12	☐ 01 ☐ 02 ☐ 03 ☐ 04 ☐ 05 ☐ 06 ☐ 07 ☐ 08 ☐ 09 ☐ 10 ☐ 11 ☐ 12	☐ 01 ☐ 02 ☐ 03 ☐ 04 ☐ 05 ☐ 06 ☐ 07 ☐ 08 ☐ 09 ☐ 10 ☐ 11 ☐ 12

		Business 1	**Business 2**	**Business 3**
27	**When did this business start operating?** 1 = LESS THAN A YEAR AGO 2 = 1 BUT LESS THAN 3 YEARS AGO 3 = 3 BUT LESS THAN 5 YEARS AGO 4 = 5 BUT LESS THAN 10 YEARS AGO 5 = 10 OR MORE YEARS AGO 6 = DON'T KNOW	☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6	☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6	☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6

28	**What was the main reason you started in this business?** 01 = INHERITED/FAMILY TRADITION 02 = UNEMPLOYED/HAVE NO ALTERNATIVE INCOME SOURCE 03 = RETRENCHED 04 = INADEQUATE INCOME FROM THE OTHER SOURCE 05 = I LIKE THE ACTIVITY 06 = I HAVE THE SKILLS FOR THIS BUSINESS 07 = I HAVE THE EQUIPMENT FOR THIS BUSINESS 08 = ACTIVITY BRINGS HIGH INCOME 09 = SMALL INVESTMENT NEEDED 10 = UNHAPPINESS WITH PREVIOUS WORK 11 = OTHER, *specify*	☐ 01 ☐ 02 ☐ 03 ☐ 04 ☐ 05 ☐ 06 ☐ 07 ☐ 08 ☐ 09 ☐ 10 ☐ 11	☐ 01 ☐ 02 ☐ 03 ☐ 04 ☐ 05 ☐ 06 ☐ 07 ☐ 08 ☐ 09 ☐ 10 ☐ 11	☐ 01 ☐ 02 ☐ 03 ☐ 04 ☐ 05 ☐ 06 ☐ 07 ☐ 08 ☐ 09 ☐ 10 ☐ 11
29	**Did you need any money to start the business?** 1 = YES 2 = NO → ***Go to Q40***	☐ 1 ☐ 2	☐ 1 ☐ 2	☐ 1 ☐ 2
30	**Did you use your own money to start the business?** 1 = YES 2 = NO	☐ 1 ☐ 2	☐ 1 ☐ 2	☐ 1 ☐ 2

		Business 1	**Business 2**	**Business 3**

31	**What was the main source of your own money?** 01 = PREVIOUS AND/OR PRESENT WAGE EMPLOYMENT 02 = OTHER BUSINESS 03 = SALE OF LIVESTOCK/CROPS 04 = SALES OF OTHER ASSETS 05 = INHERITANCE 06 = PENSION FROM WORK 07 = RETIREMENT/SEVERANCE PAY 08 = A POLICY THAT MATURED 09 = MONEY PAID OUT FROM A STOKVEL 10 = OTHER SAVINGS, *specify*	01 02 03 04 05 06 07 08 09 10	01 02 03 04 05 06 07 08 09 10	01 02 03 04 05 06 07 08 09 10
32	**What was the total amount used from your own money to start the business? (Rands)**			
33	**Did you borrow any money to start the business?** 1 = YES 2 = NO	1 2	1 2	1 2
34	**From where did you get the loan?**	YES NO	YES NO	YES NO
	1 = Loans from commercial banks	1 2	1 2	1 2
	2 = Loans from friends/relatives	1 2	1 2	1 2
	3 = Loans from credit societies/ stokvels	1 2	1 2	1 2
	4 = Loans from moneylenders/mashonisas	1 2	1 2	1 2
	5 = Loans from (business) partners	1 2	1 2	1 2
	6 = Loans from business association	1 2	1 2	1 2
	7 = Loans from NGO/CBO	1 2	1 2	1 2
	8 = Loans from others, *specify*	1 2	1 2	1 2

35	**What was the total amount of the money that you borrowed to start the business? (Rands)**	

		Business 1	Business 2	Business 3
36	**Are you presently paying off any money for the loan/s you took to start the business?** 1 = YES 2 = NO	1 2	1 2	1 2
37	**How much did you pay off in the last calendar month? (Rands)**			
38	**Did you obtain a business grant to start this business?** 1 = YES 2 = NO	1 2	1 2	1 2
39	**From where did you obtain the business grant?** 1 = From government…….. 2 = From a non-governmental organisation / CBO…… 3 = Other, *specify*	YES NO 1 2 1 2 1 2	YES NO 1 2 1 2 1 2	YES NO 1 2 1 2 1 2

General information

		Business 1	Business 2	Business 3
40	**Does the business use any raw materials, e.g. wood for furniture or steel for making gates?** *Raw material is something that you change in some way before selling it.* 1 = YES 2 = NO	1 2	1 2	1 2
41	**In the last calendar month, how much was spent on raw materials? (Rands)**			
42	**Does the business use any supplies, e.g. beverages or fruit?** *Supply is something that you <u>do not</u> change before reselling it.* 1 = YES 2 = NO	1 2	1 2	1 2
43	*In the last calendar month, how much was spent on these supplies? (Rands)*			
44	*In the last calendar month, how much money came into the business through sales or services offered, before any deductions (turnover)? (Rands)*			
45	*In the last calendar month, how much money came into the business from other sources, e.g. loans? (Rands)*			
46	**Does the business need assistance with any of the following?**	YES NO	YES NO	YES NO
	1 = Provision of an alternative site ……..	1 2	1 2	1 2
	2 = Better access to loans ……..	1 2	1 2	1 2
	3 = Assistance with marketing ……...	1 2	1 2	1 2
	4 = Better access to raw materials/supplies ……	1 2	1 2	1 2
	5 = Easing in government regulations ……..	1 2	1 2	1 2
	6 = Access to modern technology …….....	1 2	1 2	1 2
	7 = Forming contacts with others in similar businesses for cooperation…….	1 2	1 2	1 2
	8 = Other, *specify*	1 2	1 2	1 2

47	***Ask if more than one* YES *in Q46, otherwise go to Q48*** **Which of the above mentioned forms of assistance is the most important?** *Give response category number from Q46*			

People employed and labour costs

		Business 1	Business 2	Business 3
48	**Does the business employ any people, paid or unpaid, to work in this business, including household members, but excluding yourself?** 1 = YES 2 = NO	1 2	1 2	1 2
49	**How many paid and unpaid workers** (anyone working more than 1 hour per week) **were there at this business, including other household members:**	**Paid Unpaid**	**Paid Unpaid**	**Paid Unpaid**
	1 = At this time last year ……			
	2 = During the last calendar month ……			
	3 = During the last week ……			

Please note that the reference period for Q50 to Q53 is the last week

		Business 1	Business 2	Business 3
50	**In the last week, how many of these workers were:**			
	1 = Full time workers (35 hours or more per week)			
	2 = Part-time workers (less than 35 hours per week)			
51	**In the last week, how many of these workers were:**	**Paid Unpaid**	**Paid Unpaid**	**Paid Unpaid**
	1 = Male ……			
	2 = Female ……			
52	**In the last week, how many of these workers were:**	**Paid Unpaid**	**Paid Unpaid**	**Paid Unpaid**
	1 = African/Black			
	2 = Coloured ……			

	3 = Indian/Asian ……	
	4 = White ……	
	5 = Other, *specify*	

53	**In the last week, how many of these workers were:**	**Paid**			**Unpaid**		**Paid**			**Unpaid**		**Paid**			**Unpaid**			
	1 = Aged 15 – 64 years ……																	
	2 = Aged more than 64 years ……																	
	3 = Under 15 years of age ……																	

54	*What were the total wages, salaries and other benefits paid to all the employees during the last calendar month?*																	
	1 = Wages and salaries, including overtime, bonuses, etc. (Rands)																	
	2 = Payment in kind (food, clothing, drinks, etc). Give an estimated value (Rands)																	
	3 = Refund of transport costs (Rands)																	
	4 = Other, *specify*																	

Remuneration for employers and own-account workers

55	*How much was withdrawn from the business by you during the last calendar month as …*																	
	1 = Wages and salaries, including overtime, bonuses, etc for yourself? (Rands)																	
	2 = Payment in kind (food, clothing, drinks, etc)? Give an estimated value (Rands).																	
	3 = Refund of transport costs?																	

	4 = Other, *specify*	

Expenditures on fuels, materials and services

		Business 1		Business 2		Business 3	
56	**How much did the business spend on each of the following items in the last calendar month?** *State if the amount can not be separated from that of the household and give the full amount.*	**Not sepa-rated**	**Rands**	**Not sepa-rated**	**Rands**	**Not sepa-rated**	**Rands**
	01 = Electricity						
	02 = Water						
	03 = Fuel and lubricants …….						
	04 = Spare parts						
	05 = Rental of premises …….						
	06 = Rental of machinery and equipment ……..……						
	07 = Postage, telephone, printing and stationery ……..…….						
	08 = Transport of raw materials/supplies ……..……						
	09 = Repairs and						
	10 = Business services (accounting, legal, advertisement, etc)						
	11 = Licences, permits						
	12 = Interest on loans ……..……..……..……						
	13 = Repayment of loans ……..……						
	14 = Insurance premiums, mortgages/bonds……..……..						
	15 = Income tax/ levies ……..……...…….						
	16 Protection agencies……..……..……						
	17 = Other, *specify*						

Fixed capital formation

		Business 1		Business 2		Business 3	
57	**In the past year, did the business buy any of the following? If YES, what was the cost?**	**No, did not buy in the past year**	**Yes, did buy in the past year Rands**	**No, did not buy in the past year**	**Yes, did buy in the past year Rands**	**No, did not buy in the past year**	**Yes, did buy in the past year Rands**
	1 = Machinery …….……..						
	2 = Equipment and tools ……..……						
	3 = Vehicles, trailers, etc. for transporting ……...……..……						
	4 = Buildings and other structures						
	5 = Furniture …….……..……						
	6 = Other capital items, *specify*						
58	**How much money did the business make in the last calendar month after deductions (net profit)? (Rands)**						
59	**On average, how much money does the business make in a month after deductions (net profit)? (Rands)**						
60	**Does the business have any debts at present?** 1 = YES 2 = NO	1 2		1 2		1 2	
61	**How much money does the business owe at this point of time? (Rands)**						

Transport

To be answered by those offering a transport service. For others → Go to Q64

		Business 1	Business 2	Business 3
62	**What transport services does the business offer?** 1 = PASSENGER TRANSPORT WITH COMBIS/MINI-BUSES → ***Go to Q64*** 2 = PASSENGER TRANSPORT WITH CABS → ***Go to Q64*** 3 = TRANSPORT OF GOODS 4 = BOTH PASSENGER TRANSPORT AND TRANSPORT OF GOODS 5 = OTHER, *specify* _____ → ***Go to Q64***	☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5	☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5	☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5
63	**What type of goods does the business normally carry?** ***1 = SAND AND GRAVEL*** 2 = OTHER CONSTRUCTION MATERIALS 3 = GROCERIES 4 = OTHER, *specify* _____	☐ 1 ☐ 2 ☐ 3 ☐ 4	☐ 1 ☐ 2 ☐ 3 ☐ 4	☐ 1 ☐ 2 ☐ 3 ☐ 4

Now I am going to ask you about the registration of your businesse(s)

		Business 1	Business 2	Business 3
64	**You indicated that your business is not registered for VAT, why is it not registered?** ***(Repeat for each business)*** 1 = THE BUSINESS DOES NOT MEET REGISTRATION REQUIREMENTS 2 = DIDN'T KNOW THAT THE BUSINESS HAS TO BE REGISTERED 3 = OTHER, *specify*	☐ 1 ☐ 2 ☐ 3	☐ 1 ☐ 2 ☐ 3	☐ 1 ☐ 2 ☐ 3
65	***Refer to Q5 before asking this question*** **You indicated that your business is not registered for income tax, why is it not registered?** ***(Repeat for each business)*** 1 = THE BUSINESS DOES NOT MEET REGISTRATION REQUIREMENTS 2 = DIDN'T KNOW THAT THE BUSINESS HAS TO BE REGISTERED 3 = OTHER, *specify*	☐ 1 ☐ 2 ☐ 3	☐ 1 ☐ 2 ☐ 3	☐ 1 ☐ 2 ☐ 3

INTERVIEW END TIME

h	h	m	m

End of the interview.

Thank the respondent for his/her participation