Sample design considerations for informal sector surveys

Vijay VERMA
Research Professor, University of Essex Colchester CO4 3SQ, U.K. vjverma@compuserve.com

1 The survey units

An informal sector survey is concerned with economic units of special types - namely household-based and other small-scale establishments. Exactly what type of units are in the scope of the survey is a substantive issue, dependent on circumstances and data requirements, and is not considered here. However, certain common features can be noted. Typically, the survey is confined to non-agricultural units, but covers diverse activities. The units of interest lack formal or legal separation from the household as an economic entity, are owned and operated by persons alone or in partnership as self-employment activity, without employing regular employees. In developing countries in particular, the establishment is typically located in the operator's home, in temporary premises, or without a fixed place of business. In addition, the population of interest may include micro-enterprises which, while employing one or more regular employees, operate on a scale below a certain level.

In so far as the economic units of interest are household-based and have a one-to-one correspondence with households, the informal sector survey design is likely to be similar to that of a household survey. However, there are many features which distinguish such surveys from population-based household surveys. Fundamental differences arise from the fact that we are dealing with a population of units which tend to be very heterogeneous, unevenly distributed in the population, and also less stable than typical households. Furthermore, the precision requirements, and possibly even the measurement objectives, may differ from one category of units (type, size, sector of economic activity) to another. All these factors contribute to increased complexity of the survey design, including sample selection and estimation procedures. While such complexity cannot be avoided, it remains highly desirable that it be kept to a minimum. Even more important is the need to avoid complicated procedures at the implementation stage .

2 Basic design parameters

Informal sector surveys have been conducted with diverse objectives and designs in different countries. The surveys differ in scope and coverage: for instance in whether it covers only urban or all areas, only selected economic sector or all sectors, only household-based or all informal sector units, etc. Sometimes the household rather than the informal sector activity is taken as the survey unit.

A major factor affecting the design is whether, and if so how, the informal sector survey is linked to another survey, typically the labour force survey. One option is to conduct the survey simply as a module attached to the LFS. This can be economical and convenient, but has limitations. Alternatively, the survey may be conducted as a more or less separate operation, though without precluding some degree of linkage with another survey. The following discussion is mainly in the context of the latter type of situation, though many of issues raised arise more generally.

Supplementary information for survey design

The essential information for determining the sample design includes information on the number and distribution of survey units in the population, by unit characteristics. It is only on this basis that sampling rates and other aspects of the design can be determined. Normally, good information of this type is available for population-based surveys, but often that is not the case for informal sector and other surveys of small establishments. The basic information required includes: the number and distribution of informal sector units by type, economic sector, location, etc, identifying areas

of high concentration. With foresight, population and economic censuses can be (and have been) used to compile such information. Labour force surveys, provided a couple of suitable questions are included, can yield valuable information on the size and distribution of informal sector activity.

Multi-stage area-based sampling

As the population consists of numerous and scattered small units, normally no up-to-date lists exist which can be used as the sampling frame. Consequently, a multi-stage, area based design is required, just as in typical household surveys.

3 Sample size and allocation

In the context of an informal sector survey, two issues need to be discussed in particular: (i) the diversity of sampling rates required for different types of units in the same survey, and (ii) the problem of controlling the size of the sample.

In most circumstances, it is necessary to determine the required sample sizes and sampling rates separately for different sectors and types of units. Sectors can differ greatly in size, importance and data requirements. For instance, it is often necessary to over-sample the small manufacturing sector, and under-sample the diverse trading and services activities. Similarly different sampling rates are often needed for establishments of different types and sizes; for instance the number of micro-establishments, as opposed to own-account household economic units, is usually small, and the former have to be sampled at higher rates to obtain adequate numbers.

Several factors contribute to the difficulty of controlling sample size. At the overall level, the total number of units in the population is often not known. For individual areas, little or no information may be available on the number of units, particularly by sector and type.

This has consequences for all aspects of the design: stratification, type of area units used, listing within selected areas, and the final stage of sampling, as discussed below.

4 Stratification

Apart from the usual reasons for stratification, stratification has a very specific added objective in the case of an informal sector survey: *to identify and separate out different degrees of concentration of different types of units*. On this basis, different designs and sampling rates can be applied to different domains as required. Indeed, it is desirable to accommodate differences in the required sampling procedures *at the higher stages of sampling*, so that the units can be sampled in a more uniform way at the last stage of selection, which involves large and decentralised - and hence more difficult to control - operations.

The type of stratification possible depends on the information available. Such information need not be very precise or up-to-date to be useful for the purpose of stratification, so long as it is reasonably well correlated with the current relevant characteristics of the area units. Even broad classifications by geographical location, ecological characteristics, size of locality, and within larger cities by neighbourhood or district (e.g. central versus peripheral areas), etc, can be very pertinent in such surveys. Where information on population and physical size of areas is available, it is possible to introduce stratification by *population density*. Another useful stratification variable is *economic density*, i.e. the ratio of the number of informal sector units to the number of households in the area.

The potential for stratification is greatly increased when the frame contains usable information on the numbers of units by type and sector of activity. In a number of countries, population and/or economic censuses have been used to compile such information. In so far as the strata can be constructed to reflect concentrations of different types of informal sector units or sectors, sampling rates and designs for the latter can be controlled at the stratum level. Here is an example. A survey in Sri Lanka covered both agricultural and non-agricultural households, and it was required to sample non-agricultural households at a higher rate. This was achieved by over-sampling the strata (districts) which contained such households in greater concentration. This

resulted in their automatic over-representation on the average, thus reducing the need for differential selection at subsequent stages.

The following model illustrates how such stratification may be used to accommodate and control differing requirements within the same survey. Suppose that an index R_{ij} reflecting the relative concentration of sector *i* in area *j* is constructed as follows. It is the ratio of the number of informal sector units of sector *i* in the area, to the average number of such unit per area. Corresponding to a particular sector k, let K denote the set of areas (i.e. the stratum) for which R_{kj} is the largest of all R_{ij} values for area *j*. Stratum K in general will contain establishments from different sectors, and let N_{Ki} be the number of establishments of sector *i* in stratum K. The sampling rates f_K to be applied to the strata to obtain a specified number n_i of units of sector *i* is determined by the equation $\Sigma_K N_{Ki}.f_K=n_i$, summed over all strata.

To the extent a stratum 'captures' a large proportion of the establishments in the corresponding sector, the sampling requirements (such as selection rates or sample sizes) for the sector can be applied uniformly to the stratum itself, thus reducing the need to treat different sectors differently at lower stages of the sampling process.

The above example does not control separately for units of different types, such as household versus micro enterprises. It is difficult to avoid differential sampling at later stages for this purpose because of the lack of information on this aspect in most area frames. Furthermore, to control the overall sample size, it is necessary to have an idea of the relationship of the numbers of establishments (or some other measure of size) in the frame to the actual number of informal sector units expected at the time of the survey. The two numbers may be more or less highly correlated, but in general are not equal. The numbers in the frame may fall short of the actual numbers because of growth, but also because the frame may not fully cover the less visible informal sector units. Pre-testing, involving re-listing in a small sample or areas of different types may provide some information on the overall relationship between the two numbers for different types of units.

5 Selection of area units

As in the case of household surveys, the selection of area units with probabilities proportional to some measure of size (PPS) is generally a suitable method for an informal sector survey. The difference lies in what constitutes a suitable measure of size. It may be the total number of informal sector economic units, or some similar measure correlated with that number. When the survey covers a number of sectors and different types of units within sectors, a composite measure of size may be formed by giving them different weights. Variables like population density and economic density can be incorporated into the measures of size, supplementing the use of such information for stratification, as noted earlier.

Areas with low density of economic activity generally require special treatment. An important design decision concerns whether or not areas with little or no reported informal sector activity *in the frame* should be included in the sample. This can have major cost implications for the survey. If such areas are included, it is often simpler to select them with constant and relatively low probabilities rather than with the usual probability proportional to size (PPS) scheme, and then to follow that up with 'take-all' sampling at the last stage. To obtain sufficient numbers of units per area, larger PSU's may be formed by grouping neighbouring area units, or by introducing higher sampling stages.

6 Listing

The quality of estimates of aggregates (total number of economic units, employment, output etc.) from the survey depends on completeness of coverage, which in turn depends on the quality of listing. The requirement of good coverage is particularly important in the case of an informal sector survey. Visible as well as hidden units must be listed to ensure good coverage. In such a survey, the listing also has the important objectives of identifying in-scope units and other information required for stratification and selection of the units within sample areas. Furthermore, a large

number of units may have to be listed to secure sufficient sample sizes for units of different types. A 'dual approach' is desirable to ensure the coverage of all types of units. The idea of this approach is to divide the population of units into non-overlapping and exhaustive categories: (i) units which require special treatment and are appropriately listed using the establishment approach; and (ii) the bulk of smaller units which are best covered through a household listing operation. The latter can be divided into informal sector activities carried out within the household premises by persons resident in the household; and all other informal sector activities of those persons carried out without a fixed or definite location.

Listing therefore tends to be a heavy and expensive operation compared to that in a normal household survey. It involves the collection of a fair amount of substantive information over a relatively big sample.

7 Last stage of sampling

Several important aspects distinguish an informal sector survey from a typical household survey with respect to the last stage of sampling:

1. Special measures are often required to avoid the procedures becoming too complicated. Selecting units in different sectors of activity with different rates should be avoided if at all possible. It is better, as far as possible, to absorb required differences at preceding stages of sampling, which involve much smaller and better controlled operations.
2. Such complexity has often resulted in the adoption of non-probability selection procedures; this should be avoided.
3. Lack of control over sample sizes and workloads is likely to be a serious problem in informal sector surveys; consequently, the need to control these variations is typically greater than in household surveys which can more often be designed to be self-weighting.
4. It is often advisable to use the results of the listing operation to adjust the overall sampling rate, so as to control the total sample size. If M_i is the expected number and L_i the number of units actually found in area *i* after listing, then the required adjustment is to multiply all last stage rates by a uniform factor $\Sigma(M_i/L_i)$, summed over all sample areas.
5. Special treatment is required in situations when units of observation and analysis lack one-to-one correspondence with the ultimate units of sampling. Examples are the presence of several informal sector units in the same household; several types of activities carried out by the same unit; or unit with partners from different households.

REFERENCES

Verma, V. (1992). *Methods of Data Collection on the Informal Sector*. Report prepared for the International Labour Office.

RESUME

An informal sector survey is concerned with household-based and other small-scale establishments. In so far as the economic units of interest have a one-to-one correspondence with households, the informal sector survey design is likely to be similar that of a household survey. However, there are many features which distinguish such surveys from population-based household surveys. This is because the informal sector survey units are heterogeneous and unevenly distributed, and the precision requirements and measurement objectives often differ from one category of units to another. All these factors contribute to increased complexity of the survey design. The paper discusses its consequences for diverse aspects of sampling: stratification, type of area units used, listing within selected areas, and the final stage of sampling.