

Methodological issues in surveying migration across EU countries

Vijay Verma

Departimento di Economia Politica

Piazza San Francesco 8

53100 Siena, Italy

verma@unisi.it

Gianni Betti

Ilaria Vannini

1. Introduction

The freedom of movement has been one of the core rights of citizens in the EU since the 1957 Treaty of Rome. The enlargement of the European Union from EU-15 to EU-25 in 2004 and from EU-25 to EU-27 in 2007 has increased the movement and diversified the reasons for moving. Many movers which in the past were movers from outside the Union, now are movers within the Union. Not all such migrants are EU citizen. In fact, a large proportion may be migrants from outside the EU, without EU citizenship, who later move from one EU country to another. Regrettably, this freedom of movement does not fully extend to those not citizens of a EU country.

There are several sources for information concerning migration in the European Union. Some can be obtained from official registers, mainly at country level, and some from harmonised European surveys such as Labour Force Survey (EU-LFS). Survey data are useful not only for defining the migration status of each individual, but also for understanding the reasons and intentions for moving.

However, sources of information concerning migration in the European Union are not always consistent and do not cover many issues of interest.

The present paper is concerned with migration of persons living in one EU country to another EU country. While noting the usefulness of some existing EU-harmonised social surveys for assessing the social situation of people who have lived in different EU countries, our main concern is with technical issues in the development of feasible new survey instruments for the purpose: instruments targeted at movers but also permitting comparisons with non-movers.

The use of existing sources

The EU-LFS is the oldest data collection by Eurostat, first launched in 1960. A systematic common framework has been provided since 1983, updated in 1988. It is a quarterly survey with continuous fieldwork in most countries. Thanks to the large sample sizes, the high frequency and comparability across countries, the LFS can be used as a source of information for estimating basic indicators relating to migration. First of all, the variables covered in LFS include country of birth, nationality, years of residence in the present country, country of place of work; and region and country of residence one year before survey. This gives information about the magnitude and direction of migration. Secondly, it may be possible from time-to-time to add other variables concerning migration in special modules attached to the LFS, such as on "labour market situation of migrants and their immediate descendants" in LFS 2008. Finally, LFS could be used as a sampling frame or a sample for additional data collection in the form of a separate, new survey. We may also note that the Spring LFS is in many cases a larger survey, and hence provides additional information.

The European Community Household Panel (ECHP) provided an alternative though considerably smaller source. It began as a standardised survey in 1994 in EU-12, later extended to EU-15 and was concluded in 2001. ECHP data series is old, but it provides valuable information for methodological exploration for the present purpose. The survey questionnaire contains an entire section covering migration trajectories. From 2004 the survey was replaced by EU-SILC (EU Statistics on Income and Living

Conditions) covering all EU-27 and some additional countries. Unfortunately, EU-SILC contains less information about migration compared to the ECHP: it includes only country of birth and citizenship. Another potential source are the Eurobarometer surveys which cover a very wide range of topics, and some of the surveys can be designed to focus on migration-related issues. For instance, Eurobarometer 64.1 of 2005 was one such survey. It covered a wide range of mobility/migration related questions, broader than just international movement across EU national boundaries: internal migration was also covered.

Need for new instruments and approaches

It is clear that existing sources provide, at best, very limited information on the extent and characteristics of movement across EU countries. More than one instruments (questionnaires and surveys) may be required to collect the range and different types of information required: the numbers, characteristics and circumstances of migrants; reasons and conditions of the move; links with the place of origin; remittances, etc. Such information would normally be collected through household-based interview surveys. Substantive and sampling aspects of the new instruments are discussed in Sections 2 and 3 respectively. Movers across EU national boundaries have certain characteristics which makes them 'elusive' to the sampling process, meaning that it is difficult to achieve adequate samples using the normal procedures of general population sampling. Elusive characteristics and alternative sampling strategies aimed at overcoming them are discussed in Sections 4 and 5 respectively. All the methodologies presented here were developed at the request of European Commission Employment, Social Affairs and Equal Opportunities DG (Verma and Betti 2008). The authors also provided numerical illustration of the usefulness of data such as from ECHP for the construction of migration trajectories.

2 New survey instruments

The New Instrument proposed in Verma and Betti (2008) is composed of three components, differing in the scope of the information collected and the statistical units addressed: a core module covering the most basic items for the whole population; a short module addressed primarily to migrants; and a long module seeking detailed information from more recent migrants. The various modules are 'additive' in the sense that the long module would normally include most of the short module; both the long and the short modules would include the core questions.

Core module

The objective of the core module is to identify immigrants in the population, and obtain information on basic characteristics of migrants and non-migrants. The main statistical output is the estimation of proportion of immigrants in the population, and comparison of migrant and non-migrant populations in terms of basic characteristics. The core module has to be addressed to all the units in the sample, e.g. to both movers and non-movers, with the main scope of having information concerning general aspects of the entire population.

The minimum core may cover the following topics. (1) Core social variables: demographic (sex, age, marital status, household composition, country of birth and of citizenship); geographic (country, region and type of place of residence); socio-economic (employment, education, household income). (2) Questions identifying migration status, in particular pattern of migration. (3) A few additional questions for migrants only. (4) A few additional questions for non-migrants only.

Short module

The objective of the short module is to obtain information on characteristics and circumstances of migrants, including demographic characteristics, income and living conditions, employment, and main activity; on current family situation and the degree of relationship with the original (last) country; and also on the timing, reasons and conditions of the move. For each category of the main reason for moving, it would be very useful to obtain some more information in order to understand the reason more clearly and specifically. A different 'stream' of questions is asked of each respondent, depending on the main reason for the move given. The module is primarily addressed to the immigrant population, though some additional information may also be obtained for a sample of non-migrants for the purpose of comparison. This is called a 'short' module because the amount of information sought is limited – generally less than the amount which

could be reasonably collected during one interview. The interview may therefore also contain questions on other topics if desired.

Long module

The long module refers to a set of questions more detailed than the above short module. The population of interest for this more detailed questioning is recent immigrants, those coming to the country within, say, the past 5 years. The long module may include more detailed questions on reasons and conditions of the move, financial transactions with the family left behind, the effect of the move on living conditions, problems of integration in the new place, intentions to stay, and so on. More specific areas covered in the long module may include the following: satisfaction in general; satisfaction about job; problems in dealing with (making use of) the health and social protection system; comparison of the previous and the new place after migration concerning various aspects of quality of life; more detailed questions about immediate family members, such as own children who are still living abroad; language problems; cultural, religious and racial problems; loneliness, lack of social relations, and other problems of integration in the country of arrival.

This is called a 'long' module because the amount of information sought is large enough to fill the whole interview during a household survey, leaving little scope of covering additional items in the same interview.

3 Sampling aspects of household-based surveys on migration

An important new tool for obtaining information on EU migrants can be a regular household-based surveys of the general population, through which modules such as the above may be applied. It is very useful to keep in mind two different types of such household-based surveys, which differ in their objectives, or at least in the emphasis given to different types of objectives.

Survey of the general population for estimating prevalence of migration

The first type refers to a survey where the primary objective is to measure the prevalence of migration; and also variations in this prevalence by factors such as geographical location, type of place (urban-rural classification), household type and characteristics, the household's employment and income situation, and other factors associated with these variations in the propensity for individuals to move.

In measuring the proportion of migrants, the base population of interest is the total population exposed to the risk of migration. This is the same as in the case of any major survey of the general population such as the labour force survey (LFS). Hence the required structure and distribution of the two types of surveys are likely to have a lot in common. Of course, the migration survey and LFS will typically differ in design requirements and parameters: sample size and allocation, details of sample clustering and stratification, the ultimate units for which data are collected in the survey, sample rotation and other aspects of the structure of the sample over time, etc.

As to possible linkage of the migration survey to a large-scale survey of the general population such as the LFS, comprehensive but high cost stand-alone migration surveys on the hand, and much cheaper and convenient but restricted in content supplementary modules on migration, on the other, are two possible options. However, in many situations, neither of these extreme solutions – complete separation or complete integration – is the appropriate solution. Often a more practical solution is to seek partial linkage; such linkage can take many forms and can be to different degrees. Often a convenient and practical option is to draw the migration survey sample as a sub-sample from the larger general population survey.

Surveys of migrants to study conditions and consequences of migration

We have a different type of survey when the primary objective is to investigate characteristics and consequences of migration: what types of persons are involved in the migration, what are the reasons, circumstances and conditions under which they move, the circumstances and conditions of their life in the new place, and so on. The objectives may also include investigating the immediate causes and consequences of migration. The relevant base population in this type of survey is the population of migrants. What is meant is that, when the objective is to determine the conditions and consequences of migration, as distinct from its prevalence among the total population, then it is appropriate that the size and structure of the sample is

determined largely by the size and distribution of the population of migrants.

At the same time, it is important to clarify that the concept of a “survey of migrants” does not imply that the units enumerated in the survey be only migrating persons. On the contrary, it will normally be necessary in such a survey to enumerate comparable groups of non-movers, so as to provide a control group for comparison. Nevertheless, the sample size and design of a survey of migrants is determined primarily by the need to represent the population of migrants; any sample of non-migrating persons is supplementary, added to the main sample as necessary for analytical purposes.

Relationship between migration surveys with the two types of objectives

It is instructive to compare the requirements for the two types of migration surveys which have been distinguished above. Let us call them, respectively, as type “P” (for ‘prevalence’) and type “M” (for ‘migrants’) surveys. As noted, the base of type P is the general population, while that of type M is the migrant population.

A critical issue of practical importance is whether the two distinct objectives of these surveys can be satisfactorily met through a single integrated survey, or it is better to organise them as two separate – but nevertheless linked – operations, each designed to meet its own specific objectives. A related, but still a distinct, question is whether the two components can be based on the same sample of households; or whether survey M should be a subsample of survey P, of smaller size and possibly also with a different structure.

Undoubtedly, it can be more convenient and economical to cover both objectives in a single operation. An integrated “P+M” operation may well be the most suitable option under certain conditions such as the following. (1) The migrant population is not too heterogeneous or too unevenly distributed for it to be captured in a reasonable way by a general purpose sample of the population like that of survey P. (2) Survey P does not require a very large sample (which would be the case if it is required to produce estimates of the prevalence of migration for many different domains). (3) The resulting compromise sample size is not too large for the in-depth investigation which the M component typically requires. (4) Survey P itself is a stand-alone survey, so that its sample can be designed as a compromise for meeting information needs of both types of surveys. (5) But in any case, the resulting integrated interview is not too heavy to have an adverse effect on the quality (particularly completeness) of measuring the prevalence of migration which is the concern of the P component. This is a common problem which has been encountered in many other types of surveys with similarly dual objectives.

But an integrated approach may be in conflict with differing substantive and statistical requirements of the two components, with practical aspects such as respondent burden and survey costs, and especially data quality requirements. When one or more of the above conditions are violated, it is necessary to consider the possibility of operationally separating the two components.

In the case of such a separation, it would be generally appropriate to consider basing the M component on a subsample of P. The objectives of the sub-sampling would be both to reduce the sample size for survey M, and also to make it more concentrated and targeted to reflect the uneven geographical distribution of migration.

Survey at the place of origin and at the place of destination

There are two types of total populations of interest, resulting in surveys which can differ considerably in their objectives, methodology and design. One is the population at the place of origin of the migrants – from where some individuals move away and others remain behind. The second is the population at the place of destination – consisting of the original inhabitants and the migrants entering the population.

Study at the place of origin can be useful for investigating who does not move, characteristics and conditions of households experiencing migration by some of their members, characteristics and conditions of individuals experiencing out-migration by close relatives or friends. Generally such surveys cannot provide estimates of the extent of emigration, since individuals living alone before migration or entire households who have migrated cannot be captured in the sample. Our interest here is, therefore, in the surveying of migrants at the place of destination – in the country where they are currently living.

4 Elusive characteristics of the migrant populations

The populations of movers across EU national boundaries have certain characteristics which tend to make them 'elusive' to the sampling process. Developing on Kish (1991), we use the term elusive populations to mean populations for which - by virtue of their characteristics, or of the lack of suitable sampling frames, or difficulties in obtaining the required information - adequate samples cannot be defined, drawn or implemented using the normal procedures of general population sampling. Such problems arise when no frame, or at best very partial frames or lists, are available for sampling, and/or when many units are not available or willing to participate in the survey. A large proportion of the units stay away from the sampling frame, and/or from enumeration if selected for the survey. Such problems appear in a wide variety of surveys, including of migrants. Both stages - that of selecting a representative sample, and of successfully enumerating the units selected - can present severe problem and require special procedures. Addressing sampling issues requires the identification of special characteristics and circumstances of the populations to be sampled. Among these, the following are pertinent to migrants.

(1) Ill-defined population

This refers to the existence of fundamental vagueness of the definition of the population, beyond the usual problems relating to its precise demarcation in content, space and time, and the usual problems of coverage errors. Of particular importance in migration surveys is the major effect of the definition adopted as to who counts as a 'migrant'.

(2) Great heterogeneity

Migrants may be made up of diverse groups, with very different characteristics and conditions of life, in terms of histories, characteristics, current circumstances, objectives and motivations for moving, timing of moving, place of origin and destination, position and status in their new countries and so on. There are several important consequences of this heterogeneity. Most importantly, it is unlikely that a single approach or methodology will suffice to cover the diverse types of migration. Normally, a mixture of approaches will be necessary. Secondly, it is very likely that in practice it is impossible to capture all the diverse subpopulations of interest: often it will be necessary (and wise) to concentrate resources and effort over fewer, most important groups. 'Most important' does not automatically mean the largest: often smaller groups present problems of great policy concern.

(3) Small/rare population

Here the characteristic feature is that sampling the whole population with normal procedures (such as equal probability sampling of elements) does not yield a sample of adequate size for the subpopulation or domains of interest because of the small size of this subpopulation. Procedures are required for more intensive and targeted sampling for the purposes. This covers several different types of situations: rare traits; rare populations; small areas and other small domains

In most EU countries only a small proportion of the total population has lived in more than one country. For example, around 2005, fewer than 2% of EU citizens were living in another EU Member State (European Commission 2006). The average total non-national population share is larger at around 6%, but in most EU countries (with the exception of a few, mostly small countries), the majority of foreign nationals are citizens of countries outside EU-25. Hence, mostly we are dealing with small or rare population groups.

(4) Uneven distribution, patchy concentrations

This refers to the situation when the subpopulation of interest exists in large and uneven concentrations, with limited or no prior information on the location or structure of those concentrations. Typically large parts of the total population or space have very low density of the subpopulation of interest, or are even completely empty of it. A general sample design, not able to take into account the patterns of concentration, is unlikely to provide an adequate sample. Also, it cannot guarantee prior to the survey that a sample of sufficient size would be obtained for the subpopulation of interest. The unevenly distributed population of interest may not be necessarily very small in relation to the total, but being small would further increase the problems of sampling. A common characteristic of at least some groups of migrants is their clustering

together and highly uneven distribution in the host population. Special techniques such as adaptive clustering sampling are likely to prove useful. Alternatively, general spatial sampling may have to be abandoned in favour of sampling according to (possibly transient) points of concentration, using techniques such as centre sampling.

(5) *High mobility, temporal instability, changing identity and profile of units*

In any population, characteristics of unit can be expected to change over time, including births and deaths of units. With time, even the identities of certain types of units (such as households, establishments) can change. These problems require special procedures for ‘tracing’ or follow-up of the longitudinal sample, and for weighting and estimation.

The profile of persons moving across national boundaries in the EU is changing all the time. Data show that there have been a lot of changes in characteristic of migrants from Southern to Northern Europe, for example in terms of their level of education and skills. The changing patterns of movement within the EU come from legal and economic factors such as the expansion of the EU, strengthening of the common market, and rapid normalisation of the status of movers. The speed of change trends to be even more rapid at the individual level than the already rapid one at the group level. This is because of the added impact of the length of stay in the new country which operates at the individual level following a person’s migration into a new country. These rapidly changing scenarios mean that a single, static or stable, survey instrument for the collection of the required information will not suffice, and will tend to become rapidly outdated in any case. A system comprised of multiple approaches in combination which, from the outset, are designed to be adaptable to ever-changing requirements is needed.

(6) *Hidden segments of the population, difficult to identify and locate*

This is reflected in the absence, or serious incompleteness, of the sampling frame – a common problem in sampling migrant populations.

(7) *Reclusive and sensitive character of units*

‘Reclusive’ means shying away from contact with outsiders, hence in the context of a statistical survey, unwillingness to participate in it. Special procedures need to be employed to recruit and access the respondent. ‘Sensitive’ refers to nature of the information sought, or more precisely to reaction of the respondent to the enquiry. One common reason for this is the threat or shame the respondent may feel in providing the information. Special, especially indirect, procedures may be needed to encourage participation. At least a proportion of the intra-EU migrants constitute a reclusive population in this sense. two senses.

5 Alternatives of conventional household-based surveys

Conventional sampling designs uniquely link each population element to only one sampling unit. Alternative methods become necessary for obtaining adequate and efficient samples of rare, hidden, hard-to-reach, elusive populations. For such purposes, link-tracing designs often provide more practical options. In these designs, elements initially selected into the sample lead to the inclusion of other elements into the sample on the basis of some suitably defined links between the population elements. The representativeness and efficiency of the resulting design depends on the kinds of links chosen. A great variety of designs can be generated by considering different types of links. Examples are sampling with multiplicity, chain-referral methods, random-walk designs, and adaptive cluster sampling.

There are various alternatives to standard household-based surveys of migrants. The great variety of them may be grouped according to their treatment of the unit selection probabilities:

(1) Standard probability sampling procedures, but using alternative types of units, such as establishments or other institutions where migrants congregate.

(2) Methods for augmenting selection probabilities so as to increase the sample yield, such as multiplicity or network sampling, or adaptive sampling in different forms.

(3) Method requiring special procedures for assessing selection probabilities, when units are not selected directly but by employing more complex and special procedures. Examples are capture-recapture, respondent-driven and flow sampling procedures.

(4) Methods involving departures from probability sampling. These vary from less controlled haphazard and judgement sampling procedures of various types, to more controlled quota and other structured sampling procedures which try to reflect some features of probability sampling.

Probability sampling through non-household units

Large establishments

Migrants may be traced at places where they work. There are two quite distinct types of establishment-based surveys: those involving relatively large and few in number establishments, samples for which can be selected from list frames; and surveys involving numerous and small establishments, which have to be selected through area sampling. Sampling establishments from a list normally involves their selection in a single stage, which is simpler than the multi-stage sampling involved in a typical household survey. However, two aspects can make the former more complex: (1) the need to produce estimates for many different sectors (types of establishments) separately, and (2) the great variation in the size and other characteristics of the units involved. A common and efficient way is to select units with probability proportional to unit size (PPS).

Small establishments

As for small establishments, list frames are not available and usually some form of multi-stage area sampling is required. There are some special considerations which arise in the design of samples for surveys concerned with small-scale economic units, such as agricultural holdings, household enterprises, own-account businesses and other types of establishments in different sectors of the economy. Such economic units, like households, are small, numerous and widely dispersed in the population, but differ from households in being much more heterogeneous and unevenly distributed. Reference may be made to Verma et al (2010) for a practical procedure for sampling of small establishments using multi-stage designs..

Centre sampling

Many elusive populations have a tendency to congregate in particular places or territorial sites, and such centres of congregation may be used to obtain samples of the population when better alternative frames are not available. The term 'centres' is used to refer to either a partial list or a territorial place where units of interest congregate (Blangiardo, 1996) For example, many migrant groups use such centres in order to satisfy diverse needs concerning, for instance, religion, health, social interaction, recreation, shopping, and eating. Various types of sampling designs are possible, for example: taking all centres, and a sample of persons within each; taking a sample of centres, and all persons within each; or sampling both of centres, and then of individuals within selected centres (Pratesi and Rocco 2005).

Augmenting selection probabilities

Adaptive cluster sampling

In conventional sampling, the sampling design is based entirely on a priori information, and is fixed before the study begins. By contrast, in adaptive sampling, the sampling design adapts based on observations made during the survey. Among the different types of approaches under adaptive sampling the most basic and most often referred to design is the so-called adaptive cluster sampling (Thompson 1990). Adaptive cluster sampling is useful in situations where the characteristic of interest is sparsely distributed but highly concentrated, as in environment sampling or sampling of certain migrant populations. Adaptive sampling simultaneously addresses the objective of estimating the mean concentration and the objective of determining the pattern and extent of concentrations of the phenomenon of interest. It helps to concentrate resources in areas and types of units or events of the greatest interest.

Multiplicity and network sampling

The process of sample selection is applied to the sampling units and then, through counting rules linking these units to observation units, a sample of the latter is obtained. We may identify four types of links between (ultimate) sampling units and observation units (Sirken 1970): (1) one-to-one; (2) one-to-many; (3) many-to-one; and (4) many-to-many. Conventional designs are of the form (1) and (2). Rules (3) and (4) apply to designs with multiplicity. These are the multiplicity counting rules, meaning that an observation unit

is taken into the sample through linkage with a set of more than one sampling units, one or more of which have been selected into the sample. An example is the selection of migrants directly, and also through family members or colleagues.

Indirect procedures involving assessment of selection probabilities

Capture-recapture methods; mark-resighting methods

The underlying idea of capture-recapture methods is fairly simple. The size of a population can be estimated by catching and marking a sample, then releasing that sample and allowing it to mix with the other members of that population before catching a second sample. The proportion of the number in the first catch to the number in the whole population is then assumed to be the same as the proportion of marked to unmarked subjects in the second catch, thus providing an estimate of total size of the population.

In mark-resighting experiments a sample of units is marked and resightings are performed during subsequent occasions. The main advantage of the methods is that resightings are generally cheaper to acquire than physically capturing and handling the units. Resightings are also less intrusive, which is of critical importance in applications to human populations.

Respondent-driven sampling

Respondent-driven sampling is predicated on the recognition that peers (i.e. persons belonging to a given target population) are better able than researchers to locate and recruit other members of a hidden population (Heckathorn 1997). The procedure provides an ability to reach persons who shun public venues, and is particularly useful for sampling populations that do not trust the investigators. The procedure can be used for sampling groups of people who are linked by some common forms of behaviour, characteristics, situations or circumstances. Respondent-driven sampling assumes that persons at risk are connected in a network and have ties to other persons at risk.

Monitoring of flows at sample locations

This technique is used for sampling populations which are defined on the basis of some activity/state in which they engage/exist at certain specific locations. Examples are migrants passing through certain points on their route. In general, this type of sampling scheme involves the selection of locations, of observation times, of individual units, and possibly also of a subset of activities or states to be observed.

Departures from probability sampling

Despite their well-known drawbacks, non-probability sampling methods can be useful. There may be circumstances where it is not feasible, practical or theoretically sensible to do fully random sampling. Also non-probability methods are often quick, inexpensive and convenient. There can be three types of reasons for using non-probability methods for sample selection: a probability sample is considered desirable but is not feasible; a probability sample is considered unnecessary; or a non-probability sample is preferable (e.g. when the available sample size is severely restricted). A wide range of non-probability methods of sampling is encountered in practice, and the labels used to identify and distinguish them are not quite standardized. Below we categorize and briefly describe the most common methods a little more systematically.

Haphazard sampling (convenience, voluntary, snowball sampling)

Haphazard sampling means taking units into the sample without any structure, rules or fixed procedures. The basic assumption of this approach is that the population is relatively homogeneous or is well mixed, so that it does not matter much exactly how the sample is selected. Various types of sampling arrangements may be involved.

Convenience sampling. Particular units are taken into the sample simply because it is more convenient to do so than to take some other, different units. The obvious advantage is that the method is easy to use, but that advantage is offset by the great danger of bias.

Volunteer sampling. As the term implies, this type of sampling occurs when people volunteer their services for the study. Sometimes, the researcher offers payment to entice respondents. Clearly, the problem with this type of samples is that we have no evidence that the respondents are representative of the target population, and in many cases it is clear that they are not.

Snowball sampling. Snowball sampling begins with the identification of someone who meets the criteria for inclusion in the study. That person is then asked to recommend others who he/she knows, others who also meet the study criteria. The basic idea is to expand the sample originally selected by allowing the units selected to bring into the sample other units related to them in some way. Often the emphasis is on getting enough cases rather than on the units constituting a representative sample.

Judgement sampling (purposive sampling, sampling of restricted locations, of modal instances, of extremes)

The term judgement sampling is used to refer to the selection of units on the basis of the judgement that the selected units are somehow “representative” of the population (or of some specific aspects of the population) of interest. Often this method is used in exploratory studies like the pre-testing of questionnaires, in focus groups or laboratory settings where the choice of experimental subjects reflects the investigator’s pre-existing beliefs about the population. The underlying assumption is that the investigator will select units that are characteristics of the population for the purpose at hand. The critical issue here is objectivity: how much can judgment be relied upon to arrive at a typical sample?

Purposive sampling. Such a procedure may be used when the number of units to be selected is so small that variability with random selection will be excessively large, and potentially more damaging than the bias inherent in selection by judgement. The judgement is made about whether or not to include particular units into the study, rather than about some mechanism of selecting units for the sample. The assumption of the procedure is that the phenomenon of interest in the general population is represented, and can be captured, in a restricted number of carefully selected units. It is assumed that the main variability lies within units, rather than across units, from which only a very limited number has been included for study.

Restricted sampling locations. Restricted sampling locations may be used when a significant part of the population of interest is believed to be confined to a limited number of known aggregate units such as areas or establishments. This can greatly reduce survey costs, but again bias is introduced to the extent that a part of the population lying outside the areas of concentration is not covered, especially when the uncovered part has different characteristics than the part covered.

Modal instance sampling. The mode refers to the most frequently occurring value in a distribution. When we take a modal instance sample, we are trying to include the type of cases which are most frequent or typical in the population. There are number of problems with this sampling approach. Above all, how do we know what the “typical” or “modal” child worker is? Clearly, modal instance sampling is only worth considering for informal, qualitative, investigations.

Sampling for extremes. This is judgement sampling in which we look for extreme cases, for instance to understand or bring out underlying factors, causes, consequences, etc. This sort of approach can be useful for surveying extreme situations and categories affecting selective groups in the population.

Quota and other ‘structured’ sampling procedures (quota, heterogeneity, profile sampling)

This refers to forms of non-probability sampling in which the selection of units from the population is mediated through a structure or set of constraints. However, within that structure the selections are non-random. The tightness of the structure determines how close the resulting sample is to a probability sample. An example is the conventional ‘quota sampling’.

Quota sampling. This is one of the most common forms of non-probability sampling. The basic idea in quota sampling is to produce a sample matching the target population with respect to certain characteristics (e.g. age, sex) by filling quotas for each of these characteristics. It is presumed that, if the sample matches the population in these characteristics, it may also match it in the quantities we are trying to measure. The assumption of the procedure is that the main variability lies across, rather than within, the subgroups chosen, so that once sufficiently small and homogeneous groups have been defined and properly represented, it does not matter very much which particular individual units within a group are enumerated. Quota sampling is usually justified in terms of its convenience, speed and economy. It is easier to administer and has the desirable property of satisfying population proportions. Note that the method requires that good data on the

whole population be available to set the quotas. Quota sampling can be considered preferable to other forms of non-probability sampling (e.g. judgement sampling) because it forces the inclusion of members of different sub-populations. However, quota sampling disguises a potentially significant bias. Although interviewers are constrained by the quotas, they are still using some elements of judgement in the choice of the sample. The amount of flexibility interviewers have varies from survey to survey, and it is these rules and guidelines that determine how far the quota sample departs from probability-based stratified sampling.

Heterogeneity sampling. We sample for heterogeneity when the objective is to capture the full complexity of the phenomenon (e.g. diverse conditions of child labour), but representing these views or the individuals proportionately is not of concern. Another term for this is sampling for diversity. We would use some form of heterogeneity sampling when our primary interest is in getting a broad spectrum of the phenomenon, and not in identifying the “average” or “modal” conditions. This type of sampling can take the form of a ‘non-proportional quota sampling’. In this method, the minimum numbers of units to be sampled in each category are specified, without too much concern with having numbers that match the proportions in the population.

Abstract

This paper proposes procedures for the study of migration across national boundaries within the European Union. Apart from using existing EU-harmonised social surveys, new arrangements to study the phenomenon in-depth need to be developed. The scope of information to be collected may be defined in the form of a hierarchical set of modules: (1) a core module to assess numbers and basic characteristics of movers; (2) a short module to obtain information on motivations and circumstances of migrants; and (3) a long module comprising a more detailed sets of questions aimed at recent immigrants. The populations of movers across EU national boundaries have certain characteristics which makes them ‘elusive’ to the sampling process by virtue of their characteristics and lack of suitable sampling frames. These are identified, and possible sampling strategies to cope with the problem discussed.

References

- Blangiardo, G. C. (1996). Il campionamento per centri o ambienti di aggregazione nelle indagini sulla presenza straniera. Studi in onore di Giampiero Landenna, Giuffrè, Milano, pp. 13-30.
- European Commission (2006). Directorate-General for Employment, Social Affairs and Equal Opportunities. *Employment in Europe 2006*, Chapter 5 “Geographic mobility within EU”.
- Heckathorn, D. D. (1997). Respondent driven sampling: a new approach to study of hidden populations. *Social Problems* 44:174-199.
- Kish, L. (1991). Taxonomy of elusive populations. *Journal of Official Statistics*, 7(3), pp. 339-347.
- Pratesi, M., Rocco E. (2005). Two-step centre sampling for estimating the size, total and mean of elusive population. *Statistical Methods and Applications*, 14, pp. 357-374.
- Sirken, M. G. (1970). Household Surveys with Multiplicity. *Journal of the American Statistical Association*, 65, pp. 257-266.
- Thompson, S. K. (1990). Adaptive cluster sampling. *Journal of the American Statistical Association*, 85, 1050-1059.
- Verma, V., Betti, G. (2008). *Investigation into the feasibility of a new survey instrument targeted at people moving across the EU*. Final Report, Negotiated Procedure no. VT/2007/086. Brussels: European Commission Employment, Social Affairs and Equal Opportunities DG
- Verma V., Ghellini G., Betti G. (2010). Area Sampling for Small-Scale Economic Units, in Benedetti R., Bee M., Espa G., Piersimoni F. (eds.), *Agricultural Survey Methods*, John Wiley & Sons Ltd, pp. 47-73.