

# Migration Estimation Based on the Labour Force Survey: An EU-15 Perspective

Mónica Martí<sup>1</sup> and Carmen Ródenas<sup>2</sup>

<sup>1</sup> Applied Economics Department (University of Alicante, Spain), e-mail: [mmarti@ua.es](mailto:mmarti@ua.es)

<sup>2</sup> Applied Economics Department (University of Alicante, Spain), e-mail: [crodenas@ua.es](mailto:crodenas@ua.es)

## Abstract

It can be observed that the European Union Labour Force Survey (EU LFS) only allows a satisfactory estimation of the *stocks* of non-nationals or those born abroad in some countries, whereas it proves to be less than adequate in most of them with regard to migration *flows*. We believe that this very limited success is due to a two-fold statistical problem of *imprecision* and *bias*, which is intensified by the embarrassing question of *answer impossible*. These difficulties exist among the Member States to a greater or lesser degree, depending on the characteristics of the migratory domain and the particular features that the EU LFS acquires in each country.

**Keywords:** Migration, Labour Force Surveys, Sample Designs.

## Introduction

Knowing how labour migration is measured and whether comparable measurements are made throughout European Union (EU) countries are both well justified necessities. The impulse of new European requirements around employment policy creates a situation in which migration is very relevant. Reliable, accurate and harmonized statistical information is also a challenge in the context of a monetary union. However, in the EU this information is very incomplete and unequal. Lacking a specific and harmonized statistical source, member countries measure migration from the census, administrative population registers, surveys, residence permits for foreigners or, when evaluating the active foreign population, the LFS or work permits. Unsurprisingly, the lack of homogeneity among sources impedes the realisation of any kind of comparative analysis. In this context, it is important to evaluate whether the EU LFS could be an adequate option for harmonized measurement of this phenomenon.

The objective of this study<sup>1</sup> is to evaluate the quality and inter-country comparability of the EU LFS statistical information on migration. Accordingly, the first section describes its principal features and compares its migration information with that from alternative national sources, such as censuses and registers. In the second section, we investigate why there are differences in data from different sources, concentrating on how national designs of the LFS influence the estimation of migration figures. In this second part, we tackle the problems of *lack of precision*, sources of *bias* and the specific question of *answer impossible* for the estimation of migration flows. Finally, the study closes with its main conclusions and recommendations.

## 1. The EU LFS: general characteristics and alternative migration measures

The EU LFS follows a common methodology. Its questionnaire provides information on the labour situation of immigrants and their socio-economic characteristics and, in theory, its data should be

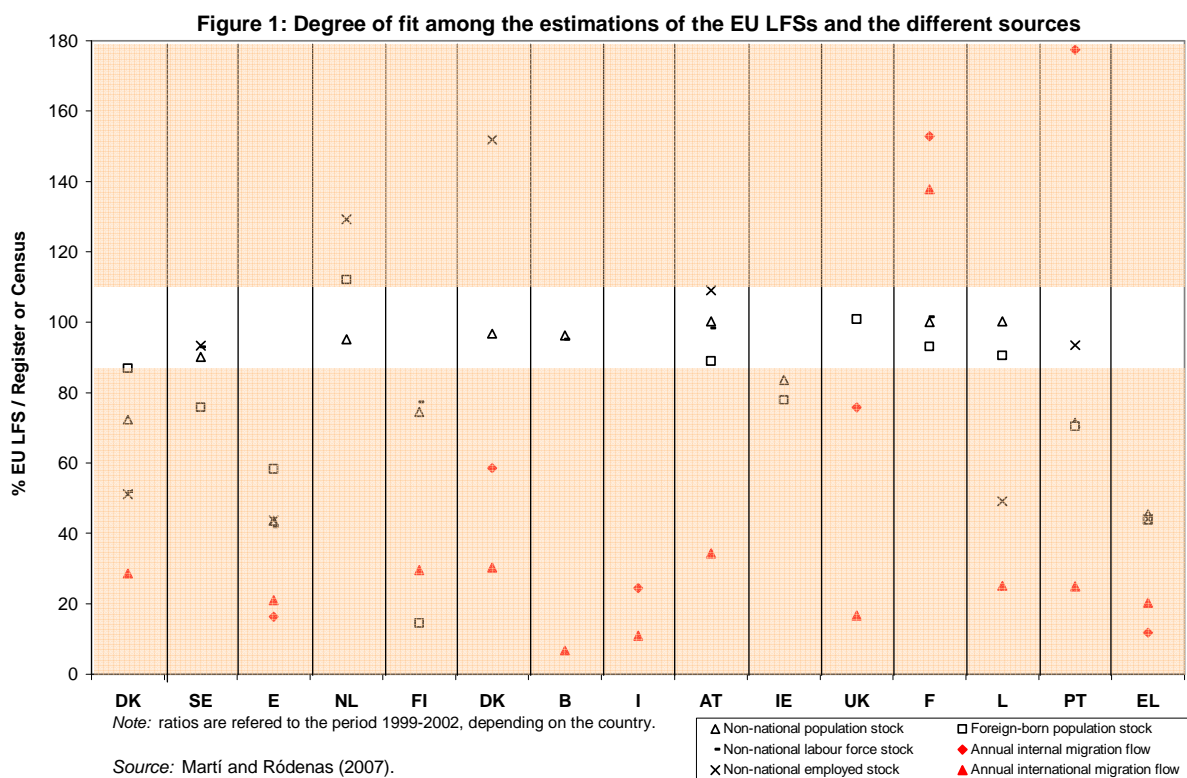
---

<sup>1</sup> This paper is a resume. For the complete article see Martí and Ródenas (2007).

comparable between EU countries. However, the National Statistics Offices are responsible for designing their own surveys -subject to restrictions imposed by community regulations-, selecting the population sample to be surveyed in the EU LFS, conducting the interviews and sending the data to Eurostat.

One method of approaching the analysis of the quality of the EU LFS migration data is to compare it with that of other national sources, such as censuses or population registers. In theory, there should be a certain degree of coincidence among different sources, although we cannot expect absolute similarity. The EU LFS and censuses measure flows retrospectively; they compare current place of residence with that of a previous date and, therefore, capture an immigrant in cases where the current place of residence does not coincide with that of the earlier date. Registers are made from information provided by the population and are, therefore, constantly updated.

In figure 1, we show the relative values of six dimensions of the migration phenomenon for the EU-15. These are annual internal and international mobility *flows*; *stock* of foreigners defined through nationality and place of birth, and finally *stock* of the non-national labour force and the employed non-national stock. Each of these six variables has been taken from the appropriate national population census or register and has been compared with the corresponding value estimated by the EU LFS.



If we admit differences of  $\pm 10\%$  among the statistical sources, the LFS estimations of *stocks* would be a little more accurate than those of *flows*. In fact, the estimation of migration *flows* through the LFS presents a high level of discrepancy in all cases when compared to information from registers or censuses, both for internal and international *flows*. In general, the LFS substantially underestimates annual *flows*, except in the cases of Austria, France and Portugal in which it overestimates. However, it performs better when estimating *stocks* and, in some countries – Austria, Belgium, France, Luxembourg, Sweden or the United Kingdom- the degree of coincidence is very high.

These results raise two very clear questions: i) why does the LFS estimate some variables better than others? and ii) why do some countries have higher coincidence than others when considering different sources? Our main hypothesis is that this situation is the result of the specific national sample design of each country. In fact, although the EU LFS may be homogeneous and Eurostat may have minimum requirements in terms of sample error, the design of the LFS is not

identical in all countries. This, contrary to expectations, is very important in order to produce correct and comparable results among different countries, especially in certain *items*, such as those related to mobility. In theory, the EU LFS appears to be harmonized. However, the national differences in sampling frame, sample stratification, rotation pattern, final sample unit and, among others, domain size impede the correct and harmonized collection of data.

## 2. How the National Design of the LFS Influences the Estimation of Migration

### 2.1. The Accuracy of Estimations

The accuracy of an estimation depends on the size of the sample in relation to the domain size, on the heterogeneity of the variable studied in the final sampling unit, and on the efficiency of the stratification technique. With regard to the first aspect, which, in this case, is mobility in the total population, it is known that samples sized to make general conclusions or to measure particular parameters cannot be used to analyse very infrequent characteristics. The permissible error could be substantially increased as the number of interviews of people with this feature would not be sufficient.

In the case of the EU-15 countries, the migration domain sizes examined are generally reduced and unequal. With regard to *stocks*, with the exception of Luxembourg with values of above 30%, non-national proportions are around 5% of the national population. The *stock* of the foreign born population is about 8% and, finally, the *stock* of active and/or employed non-nationals is between 2% and 4% of the national population, also with the exception of Luxembourg.

Migration domain size is even smaller in the case of *flows*, both international and internal. Annual immigration from abroad does not rise above the 3% of Luxembourg and is only 0.3% in Finland. With regard to internal migration, differences among countries are also very significant: more than 6% of the population annually change their municipal residence in Greece, Ireland, Sweden and the UK, whereas this proportion is little more than 2% in Italy and Spain. In general, the domain sizes for those countries for which we have information are small. This is very important when evaluating the quality of statistical migration information from the LFS.

Clearly, the domain size studied is relevant for the wellness of estimations obtained through surveys. As shown by Purcell and Kish (1979:367), when a domain does not reach a sufficient size, we cannot always guarantee a satisfactory application of traditional sampling. This does not only happen with the so called *mini domains*, where between 0.01% and 1% of the reference population presents the characteristic studied, which is precisely the case of annual international immigration in many countries. We can also encounter problems with a *minor domain*, in which the characteristic is found in between 1% and 10% of the reference population, which is the case in almost all countries for the remaining migration dimensions.

Evidently, for the same sampling fraction, a domain or characteristic that is more frequent in one country than in another will be better captured in the former and, in addition, a low frequency characteristic will be better estimated with a larger sampling fraction. As we can see in [table 1](#), in Austria, Belgium, Greece, Ireland, Luxembourg and Portugal, more than 0.5% of the population are sampled, whereas in France, the Netherlands, Finland and the UK, the sampling fraction is below 0.25%. However, large domains and sampling fractions do not always lead to estimations of sufficient quality; in Greece and Ireland, with similar characteristics, well-fitting estimations are not obtained. Additionally, countries such as Germany, France, the Netherlands, Sweden and the UK have acceptable estimations for relatively large domains but have small sampling fractions.

It is possible that the high level of error found in the estimation of the migration domain could be even higher because of the fact that most countries design their LFS through cluster sampling, in which the final sample units are households or dwellings, which include one or more individuals. For cluster sampling to be as precise as simple random sampling for a certain characteristic there should be no correlation in the variable among the members of the cluster, in other words, no homogeneity in the variable for all the members of the household.

It is fairly common for the migration characteristic to affect the whole family group, especially in countries with strong Catholic roots. As demonstrated in Spain by Martí and Ródenas (2004), this means that when a sample unit (hereafter SU) is interviewed after migration, in general there is not only one migrant captured but the whole household. Therefore, the larger the household sizes, the greater the number of migrants counted. The same applies the other way round; when there are problems in capturing immigrants –for example, due to domain size, non-response or sample loss- each non-captured SU results in far fewer individual migrants being counted.

With a domain of these characteristics (reduced size and homogeneous in the final sampling unit), stratified sampling is the technique used to obtain trustworthy estimations, provided it is possible to find a population partition consisting mainly of migrants. Stratification, insofar as in each stratum it is possible to group units that are homogenous to each other and heterogeneous in relation to the other strata, reduces the variance of estimators, increases precision and, finally contributes to reducing sampling errors. For this method to be efficient, the variables used for the stratification should be correlated with the object variables of the study. In the majority of EU countries, the stratification criterion used is geographical. If the immigrants in one of these countries are concentrated in a certain region, this procedure could perhaps reduce sampling errors, but if this is not the case (which is most likely), sampling errors will still be high. Sweden is the only country that uses the nationality variable as a stratification criterion –see [table 1-](#), meaning that it is possible to reduce sampling error in the estimation of the *stock* of foreign immigrants. In fact, the estimation of the *stock* of non-nationals in this country fits fairly well with its registry value.

## **2.2. Bias in the estimations**

Apart from accuracy related problems deriving from sample size and domain characteristics, the estimation of the migration phenomenon is also affected by the typical sources of bias, which in this particular case could have important effects. We are referring to the suitability of the sampling frame, its updating and to non-response.

With regard to the sampling frame suitability, it can be seen in [table 1](#) that more than half of the countries only use private households to construct their samples and do not sample collective households. This sampling plan characteristic could be a source of underestimation. In the case of foreign immigration flows, the first places of residence for many foreign immigrants are reception centres, hostels or similar establishments [OECD (2003:5)] and, in the case of internal migration flows, many high migration groups such as military personnel or students are left out of the surveys [Rees and Kupiszewski (1999: 562)].

A fundamental requirement for good estimations from periodic surveys is the frequent updating of the sampling frame. Accordingly, it could be thought that samples based on continually updated population registers would probably obtain better results than those based on population censuses, which are usually made every ten years. Belgium, Denmark, Italy, Luxembourg, Finland and Sweden base their samples on population registers. The Netherlands and the UK use a sampling frame based on addresses supplied by the postal service, which are continually updated. However, Austria, Germany, Greece, Spain, France, Ireland and Portugal design their samples from censuses, which could, in principle, be affected by this problem –see [table 1-](#).

In countries in which the sample is census based, demographic growth areas due to migration could go unaccounted for until the next census, and it is also possible that in the meantime the flow may have reduced or that the migrants may have moved to other places. Therefore, when samples are based on infrequently updated sampling frame, it will be more difficult to capture migration movements as and where they happen. The solution to the problem lies in assiduous updating of the sampling frame. Austria and Germany annually update their census information and Spain and Ireland update quarterly, thus minimising the problem.

In any case, the most important source of bias in the LFS comes from non-response, either from there being no people resident in the household or due to reluctance or refusal to cooperate

with the survey. Although non-response is a non-sampling error that is difficult to quantify, in [table 1](#) we show the non-response percentages of thirteen EU countries.

Obviously, in countries where participation in the survey is obligatory – such as Belgium, Italy or Portugal- there is higher participation. Chief among the countries with high levels of non-response (over 15%) are Denmark, Luxembourg, the Netherlands and the UK. Non-response in itself is not what generates bias; this happens when it is correlated with population characteristics. If this correlation is not adequately corrected, bias is produced in the estimation of these characteristics. There are various indications that migration is probably correlated to non-response. Some studies<sup>2</sup> show that migrants, as opposed to non-migrants, have a greater probability of forming part of the non-response group; and that this applies especially to single migrants. In Denmark, analysis of non-response patterns shows that there is a high level among foreigners<sup>3</sup>.

The re-weighting or post-stratification procedure serves to correct the bias brought about by non-response. Correction bias requires the use of auxiliary variables related to the migration phenomenon in order to re-establish the weight or representation of the phenomenon in the population as a whole. The use of this technique is naturally subject to the availability in each country of a statistical source of control that provides updated auxiliary information. As shown in [table 1](#), there are no countries working with variables related to internal migration and only four of the fifteen –Austria, Germany, the Netherlands and Luxembourg- introduce the nationality variable into their post-stratification criteria. These are the only countries where bias in the estimation of the *stock* of foreigners is corrected. In fact, in these four countries the estimation of the *stock* of foreigners through the nationality variable coincides with the value given by censuses and registers.

### **2.3. The Problem of *Answer Impossible*: The Question to Capture Migration Flows and the Sample Rotation Pattern**

In the earlier sections, we refer to problems common to the estimation of any domain, but when we estimate the migration phenomenon in terms of flow, there is an additional problem that increases imprecision (by reducing sample size) and inevitably undervalues all migration estimations if no correction factors are introduced. This problem comes from the question designed to calculate migration *flows* in the LFS: “*what was your place of residence one year ago?*” The formulation of this *item*, with a temporal limit of one year, to a sample in which the rotation scheme divides it into sub-samples with different participation durations, leads to part of the sample never being able to give a positive response as they have been in the sample and, by definition, the same place of residence, for over one year.

The frequency of the EU LFS and the refreshment patterns of the surveyed samples are controlled by community regulations. In cases where there is still only one annual LFS, it must be made in spring and at least a quarter of the SU must come from the previous year’s survey. In countries with monthly or quarterly LFSs, community legislation makes no obligations with regard to the determination of rotation schemes. Hence, the dynamics of the sample panel are decided by each country. This means that in annual LFSs, by law the individuals of 25% of the sample households can never answer that their country/region of residence one year ago was different to the present; this possibility is incompatible with their survival in the sample. This is only a minimum percentage so there is nothing to stop countries from establishing higher percentages and repetition rates, as shown in columns (1) and (2) of [figure 2](#).

From column (2) it can be seen that the proportion of the SU that is repeated a year later varies from the minimum in Belgium with no repetition, to the 75% of Germany. In the case of countries with annual LFS sample refreshment, this percentage coincides exactly with that of the SU with *answer impossible* –see column (3)-; i.e. with those people that, having remained in the sample over one year, could not have migrated in the previous twelve months, as in Germany and

---

<sup>2</sup> CSO and ONS (1998:23) or Clark et al. (1998:8).

<sup>3</sup> Denmarks Statistik: *Internal Quality report standard form for LFS data* (p. 11).

Luxembourg. However, for the countries that have established quarterly sample rotation schemes, the calculation of the percentages of column (3) is more complex.

**Figure 2. Sample replacement and migrants not counted by country (LFSs 2004-2005)**

EU-15 Countries	Percentage and frequency of sample replacement (1)	Percentage of sample not replaced annually (2)	Answer impossible rate (3)
<b>Austria (AT)</b>	12.5% quarterly	50.0%	68.75%
<b>Belgium (B)</b>	50.0% quarterly	0.0%	0.0%
<b>Denmark (DK)</b>	33.3% quarterly	33.3%	41.7%
<b>Germany (D)</b>	25.0% annually	75.0%	75.0%
<b>Greece (EL)</b>	16.7% quarterly	33.3%	58.3%
<b>Spain (E)</b>	16.7% quarterly	33.3%	58.3%
<b>France (F)</b>	16.7% quarterly	33.3%	58.3%
<b>Ireland (IE)</b>	20.0% quarterly	20.0%	50.0%
<b>Italy (I)</b>	25.0% quarterly	50.0%	56.25%
<b>Luxembourg (L)</b>	75.0% annually	25.0%	25.0%
<b>Netherlands (NL)</b>	na	na	na
<b>Portugal (PT)</b>	16.7% quarterly	33.3%	58.3%
<b>Finland (FI)</b>	20.0% quarterly	40.0%	55.0%
<b>Sweden (SE)</b>	12.5% quarterly	50.0%	68.75%
<b>United Kingdom (UK)</b>	20.0% quarterly	20.0%	50.0%

Source: Table 1 and Martí and Ródenas (2007).

In national surveys designed with rotating sample panels, an SU remains in the sample over a limited number of interviews before being substituted. In these cases, a predetermined percentage of the sample is replaced every quarter after each interview –see column (1) of [figure 2](#)-. For this, the full sample is divided into sub-samples or waves. For example, suppose that it is established that an SU is interviewed on only six occasions, this means that the sample is divided into six and after each interview, one sixth are replaced by new units to refresh the sample. In any given period, a sixth of the sample is interviewed for the final time, another sixth for the fifth time, etc. In general, when the retrospective question is made (in moment  $t$ ), the sample is divided into “ $n$ ” sub-samples or waves with a certain number of immigrants in each one.

Under some assumptions<sup>4</sup> we can divide the immigrants of each wave into four sub-groups according to the quarter in which migration took place. Each  $M_{ij}$  represents the immigrants of wave “ $i$ ”, with “ $j$ ” quarters from when they migrated to when they were captured by the survey and made their first interview. If we had *a priori* knowledge of all the sample immigrants of year  $t$  in the second quarter<sup>5</sup>,  $t(2)$ , they could be distributed as in [figure 3](#) when the interviews are conducted in consecutive quarters; and as in [figure 4](#) when they alternate with rest periods, as is the case in Denmark, Italy and Finland.

In [figure 3](#), we simultaneously show three different sample designs: that of Austria and Sweden, of Greece, Spain, France and Portugal and, finally, that of Ireland and the UK. In Austria and Sweden, the sample is divided into eight waves and in each quarter an eighth of the sample is replaced; in Greece, Spain, France and Portugal, there are six waves, with a sixth of the sample being replaced every quarter; and in Ireland and the UK, there are five waves, meaning that every quarter a fifth of the sample is replaced.

<sup>4</sup> See Ródenas and Martí (1997).

<sup>5</sup> On the second quarter of year the sample is questioned about their place of residence exactly one year ago, as Eurostat requires [see Council Regulation (EC) n°577/98].

**Figure 3. Migrants by wave and quarter in which they move**

Year (quarter)	Waves								
t-3(3)					Austria and Sweden			M <sub>8,4</sub>	
t-3(4)							M <sub>7,4</sub>	M <sub>8,3</sub>	
t-2(1)	Ireland and United Kingdom						M <sub>6,4</sub>	M <sub>7,3</sub>	M <sub>8,2</sub>
t-2(2)					M <sub>5,4</sub>		M <sub>6,3</sub>	M <sub>7,2</sub>	M <sub>8,1</sub>
t-2(3)				M <sub>4,4</sub>	M <sub>5,3</sub>		M <sub>6,2</sub>	M <sub>7,1</sub>	W <sub>8</sub>
t-2(4)			M <sub>3,4</sub>	M <sub>4,3</sub>	M <sub>5,2</sub>		M <sub>6,1</sub>	W <sub>7</sub>	W <sub>8</sub>
t-1(1)		M <sub>2,4</sub>	M <sub>3,3</sub>	M <sub>4,2</sub>	M <sub>5,1</sub>		W <sub>6</sub>	W <sub>7</sub>	W <sub>8</sub>
t-1(2)	M <sub>1,4</sub>	M <sub>2,3</sub>	M <sub>3,2</sub>	M <sub>4,1</sub>	W <sub>5</sub>		W <sub>6</sub>	W <sub>7</sub>	W <sub>8</sub>
t-1(3)	M <sub>1,3</sub>	M <sub>2,2</sub>	M <sub>3,1</sub>	W <sub>4</sub>	W <sub>5</sub>		W <sub>6</sub>	W <sub>7</sub>	W <sub>8</sub>
t-1(4)	M <sub>1,2</sub>	M <sub>2,1</sub>	W <sub>3</sub>	W <sub>4</sub>	W <sub>5</sub>		W <sub>6</sub>	W <sub>7</sub>	W <sub>8</sub>
t (1)	M <sub>1,1</sub>	W <sub>2</sub>	W <sub>3</sub>	W <sub>4</sub>	W <sub>5</sub>		W <sub>6</sub>	W <sub>7</sub>	W <sub>8</sub>
t (2)	W <sub>1</sub>	W <sub>2</sub>	W <sub>3</sub>	W <sub>4</sub>	W <sub>5</sub>		W <sub>6</sub>	W <sub>7</sub>	W <sub>8</sub>

Note: ■ Subgroups of migrants counted ■ Subgroups of migrants not counted

Hence, if on the first day of the second quarter of year t respondents are questioned about their place of residence exactly one year previous, the quantity  $M^{t(2)}_{no\ count}$  would not be counted as immigrants:

$$M^{t(2)}_{no\ count} = M_{2,4} + \sum_{j=3}^4 M_{3,j} + \sum_{j=2}^4 M_{4,j} + \sum_{i=5}^n \sum_{j=1}^4 M_{i,j}$$

and as immigrants there would be the figure  $M^{t(2)}_{count}$ :

$$M^{t(2)}_{count} = \sum_{j=1}^4 M_{1,j} + \sum_{j=1}^3 M_{2,j} + \sum_{j=1}^2 M_{3,j} + M_{4,1}$$

In this way, of the thirty two initial immigrant subgroups in Austria and Sweden, only ten are captured and the remaining twenty two will never be counted as immigrants in moment t(2), which, under our original assumptions, represents 68.75% of the sample immigrants. En Greece, Spain, France and Portugal, of the twenty four subgroups, fourteen would not be counted (58.33%); and in Ireland and the UK, half of the twenty subgroups (50%) are not captured. As we can see, the larger the number of waves, the greater the percentage of sample individuals that can never answer that their place of residence one year previous was different to the present.

**Figure 4. Migrants by wave and quarter in which they move**

Year (quarter)	Denmark			Italy				Finland					
	Waves	Waves	Waves	Waves	Waves	Waves	Waves	Waves	Waves	Waves	Waves		
t-3(3)													
t-3(4)												M <sub>5,4</sub>	
t-2(1)			M <sub>3,4</sub>				M <sub>4,4</sub>				M <sub>4,4</sub>	M <sub>5,3</sub>	
t-2(2)			M <sub>3,3</sub>			M <sub>3,4</sub>	M <sub>4,3</sub>				M <sub>4,3</sub>	M <sub>5,2</sub>	
t-2(3)			M <sub>3,2</sub>			M <sub>3,3</sub>	M <sub>4,2</sub>				M <sub>4,2</sub>	M <sub>5,1</sub>	
t-2(4)			M <sub>3,1</sub>			M <sub>3,2</sub>	M <sub>4,1</sub>				M <sub>3,4</sub>	M <sub>4,1</sub>	W <sub>5</sub>
t-1(1)		M <sub>2,4</sub>	W <sub>3</sub>		M <sub>2,4</sub>	M <sub>3,1</sub>	W <sub>4</sub>		M <sub>2,4</sub>	M <sub>3,3</sub>	W <sub>4</sub>	W <sub>5</sub>	
t-1(2)	M <sub>1,4</sub>	M <sub>2,3</sub>	W <sub>3</sub>	M <sub>1,4</sub>	M <sub>2,3</sub>	W <sub>3</sub>	W <sub>4</sub>	M <sub>1,4</sub>	M <sub>2,3</sub>	M <sub>3,2</sub>	W <sub>4</sub>	W <sub>5</sub>	
t-1(3)	M <sub>1,3</sub>	M <sub>2,2</sub>		M <sub>1,3</sub>	M <sub>2,2</sub>	W <sub>3</sub>		M <sub>1,3</sub>	M <sub>2,2</sub>	M <sub>3,1</sub>	W <sub>4</sub>		
t-1(4)	M <sub>1,2</sub>	M <sub>2,1</sub>		M <sub>1,2</sub>	M <sub>2,1</sub>			M <sub>1,2</sub>	M <sub>2,1</sub>	W <sub>3</sub>			
t (1)	M <sub>1,1</sub>	W <sub>2</sub>		M <sub>1,1</sub>	W <sub>2</sub>		W <sub>4</sub>	M <sub>1,1</sub>	W <sub>2</sub>	W <sub>3</sub>		W <sub>5</sub>	
t (2)	W <sub>1</sub>	W <sub>2</sub>	W <sub>3</sub>	W <sub>1</sub>	W <sub>2</sub>	W <sub>3</sub>	W <sub>4</sub>	W <sub>1</sub>	W <sub>2</sub>	W <sub>3</sub>	W <sub>4</sub>	W <sub>5</sub>	

Note: ■ Subgroups of migrants counted ■ Subgroups of migrants not counted

In Denmark, Italy and Finland, after various interviews, the SUs are temporarily substituted and are later reincorporated into the sample. To be precise, in Denmark, the sample is divided into three waves, each being interviewed over two consecutive quarters and later after one year. In Italy and Finland, the sample is divided into four and five waves respectively, with rest periods of two quarters between the second and third surveys. Therefore, in Denmark, for example, in moment  $t(2)$  –see [figure 4](#)– a third of the sample is interviewed for the first time, another third for the second time and the remaining third, after having had a one-year rest period, are interviewed for the third and final time.

In the case of Denmark, of a total of twelve subgroups of immigrants (three waves by four quarters), five can never answer that their place of residence was different one year previous (41.7% of the sample immigrants); in the case of Italy, nine of the sixteen subgroups (56.25%) and finally, in Finland eleven from a total of twenty (55% of the sample immigrants). If we compare the *answer impossible* percentages of Finland and the UK, both with five LFS waves, we observe that the Finnish sampling pattern, which includes rest periods, raises the percentage even further.

As with the case of the number of waves affecting the *answer impossible* percentage—more waves leading to a higher percentage—, the inclusion of rest quarters after interviews also affects the proportion of the sample that can never give a positive answer to the migration question. Moreover, also relevant is the moment at which a wave takes its rest period. Rest periods immediately after the first interview are enormously detrimental to migration estimation, and that the *answer impossible* rate grows with the number of rest quarters. The negative effects of rest periods decrease the further away they move from the first interview, in such a way that after the fourth or subsequent interview, introducing rest periods has no additional effects on the *answer impossible* rate. This is obviously the case because there can no longer be any more timing problems with the migration *item*. Compared to continuous sampling patterns, the additional *answer impossible* cases created by incorporating three or more rest quarters after the first interview raise the percentage from 37.5% to 75% in cases with four waves; and from 68.5% to 87% in cases with eight waves.

### 3. Conclusions and main recommendations

The objective of the study is to evaluate whether the current EU LFS could be an alternative statistical source to national censuses and registers for the harmonised measurement of migration. After analysing consistency among national data from various sources, we conclude that the degree of coincidence among them is only acceptable in some countries when measuring the migration variable in terms of *stocks*, but not for migration *flows*.

We think that this limited success is due to a two-fold statistical problem of lack of precision and of bias. These two serious hindrances are present throughout the Member States to a greater or lesser extent according to the characteristics of the migration domain and the particular features of the LFS in each country. The size of the migration domain in terms of *flows* –which, except for internal migration in the UK and Sweden, does not rise above seven percent of the population– results in it being a minor or mini-domain. In these conditions –the domain does not reach a sufficient size –we cannot guarantee the satisfactory application of traditional sampling, which is what happens in the European case. Therefore, we believe that part of the lack of precision comes from the combination of reduced domain size and the relatively small sampling fractions of the national LFSs. This imprecision is also brought about by the homogenous distribution of the migratory characteristic in households, especially in predominately Catholic countries.

However, if the root of the problem were only lack of precision, we would expect that by either taking various years for the same country or by taking various countries for the same year, we would find that on some occasions the LFS data would coincide with that of censuses and/or registers. As this is not verified, we have to add the problem of bias in the LFS estimations of migration.

The updating of the sampling frame and the inclusion or not of collective households in the survey are also elements that bring about bias in measuring migration, both in terms of *stocks* and *flows*. But, there is no doubt that non-response is the main source of bias. There are few EU countries in which participation in the LFS is compulsory, which means that in the rest, the percentage of the sample presenting non-response is sufficiently large to think that it could generate problems of bias in the estimation of migration, as it is positively correlated with this characteristic. As this bias is not adequately corrected by the different national re-weighting or post-stratification procedures, the LFS estimations are generally far removed from those of censuses and registers. However, in countries that post-stratify through the nationality variable (Austria, Luxembourg, Germany and the Netherlands), the estimation of *stocks* fits reasonably well, even in the Netherlands, where non-response is very high.

Finally, it has been shown that the migration question and the sample rotation pattern generate high *answer impossible* rates, which has repercussions on both precision and bias in the estimations of migration *flows*. Apart from the number of waves affecting the *answer impossible* rate –more waves leading to a higher percentage–, the inclusion of rest quarters after some interviews also affects the proportion of the sample that can never give an affirmative answer to the migration question. In the extreme case, when there are rest periods after the first interview, up to 87% of the respondents –in cases of eight waves– cannot answer the migration question. A simple way to solve this problem is to only ask questions relative to the previous year on the first wave of the survey, and subsequently apply specific evaluation factors as with the Spanish LFS.

In summary, we consider that the set of factors that differentiate the design of the national LFSs –such as sampling frame, sampling fractions, sample rotation patterns or stratification criteria–, do not support the conclusion that migration data from this source is harmonised. We should also bear in mind the differences that both domain size and homogeneity in the final sampling unit can generate in each country. We think that a greater amount of common practice in the procedures of the EU LFSs is worthless unless urgent attention is given to the serious bias problems generated by non-response and its subsequent treatment.

## References

- CSO and ONS (1998) Indicators of Migration between the Republic of Ireland and The United Kingdom, *Eurostat Working Papers*, E3/1998-1.
- Clark, J. et al (1998) Documentation of Eurostat's Database on International Migration: Labour Data, *Eurostat Working Paper* E3/1998-16.
- Council Regulation (EC) n°577/98 of 9 March 1998 on the *Organisation of the EU LFS*.
- Denmarks Statistik (internal document) *Internal Quality Report Standard Form for LFS Data*.
- Eurostat (2007) Labour force survey in the EU, Candidate and EFTA Countries. Main characteristics of the national surveys 2005, *Methodologies and Working papers*.
- Martí, M. and Ródenas C. (2004) Migrantes y migraciones: de nuevo la divergencia en las fuentes estadísticas, *Estadística Española*, 156 (2º term), 293-321.
- Martí, M. and Ródenas C. (2007) Migration Estimation based on the Labour Force Survey: An EU-15 Perspective, *International Migration Review*, 41(1), 101-126.
- OECD (2003) *Sources and Comparability of Migration Statistics*, Statistical Annex, ([http://www.oecd.org/document/36/0,2340,en\\_2825\\_494553\\_2515108\\_1\\_1\\_1\\_1,00.html](http://www.oecd.org/document/36/0,2340,en_2825_494553_2515108_1_1_1_1,00.html))
- Purcell, N. and Kish L. (1979) Estimation for Small Domains, *Biometrics*, 35, 365-384.
- Rees, P. and Kupiszewski M. (1999) Internal Migration: What Data are Available in Europe?, *Journal of Official Statistics*, 15(4), 551-586.
- Ródenas, C. and Martí M. (1997) ¿Son bajos los flujos migratorios en España?, *Revista de Economía Aplicada*, V(15), 155-171.

**Table 1. Technical features of the national European Union Labour Force Surveys (2001-2003)**

	Austria	Belgium	Denmark	Germany	Greece	Spain	France	Ireland	Italy	Luxemb.	NL	Portugal	Finland	Sweden	UK
<b>Frequency of results</b>	Quarterly	Quarterly	Quarterly	Annual	Quarterly	Quarterly	Quarterly	Quarterly	Quarterly	Annual	Monthly	Quarterly	Monthly	Monthly	Quarterly
<b>Sampling frame</b>															
<i>Basis of sampling frame</i>	Austrian ousing Census	Population register	The Population Register and The Unempl. Register	“Old” Länder: Popul.Census and Census of buildings and housing of 1987. “New” Länder: Popul.register	Population Census	Population Census	Population Census	Population Census	Municipal Population Register	Central Population Register	Geogr. Register (addresses compiled by the Post Office) and Register of houses in Amsterdam	Population Census	Central Population Register	Statistics Sweden’s register of the Total Population (RTB) and Sweden’s Empl. Register	Most of Great Britain: Postcode Address File. North of the Caledonian: Telephone directory. Northern Ireland: the Rating and Valuation Lists
<i>Updating of the basis</i>	Annual	n.a.	n.a.	Annual	n.a.	Quarterly	n.a.	Quarterly	n.a.	n.a.	n.a.	n.a.	Monthly	Daily	n.a.
<i>Lowest level SU</i>	Dwelling	Household	Person	Clusters of households	Household	Dwelling	Dwelling	Household	Household	Household	Household	Household	Person	Person	Household
<i>Collective househ. sampled?</i>	No	No	Yes	Yes	No	No	Yes	No	No	No	No	No	Yes	Yes	Yes
<i>Criteria for stratification</i>	Region and socioeconomic	Region (exists at province level. NUT-II)	Registered unemployment	Region	Region	Region and socioeconomic	Region	Region	Region	None	Region	Region	Region and demographic (sex and age band)	Region, sex, age, nationality and employment	Region
<b>Sample design</b>															
<i>Sample size</i>	31,000 dwellings	48,000 households	16,665 people	150,000 households	30,000 households	65,208 dwellings	54,000 dwellings	39,000 households	75,516 households	8,500 households	10,000 households	20,000 households	12,000 people	59,400 people	57,000 households
<i>Sampling fraction</i>	0.7%	1.11%	0.4%	0.45%	0.87%	n.a.	0.17%	3.3%	0.36%	5%	0.07%	0.68%	0.2%	n.a.	0.2%
<i>Rotation scheme</i>	8-Waves	2-Waves	3-Waves (2-3-1)	4-Waves	6-Waves	6-Waves	6-Waves	5-Waves	4-waves (2-2-2)	None	4-Waves	6-Waves	5-Waves (3-2-2)	8-Waves	5-Waves
<i>% of the sample being replaced each year</i>	50% One eighth each quarter	100% One half each quarter	66.67% One third each quarter	25% One fourth each year	66.67% One sixth each quarter	66.67% One sixth each quarter	66.67% One sixth each quarter	80% One fifth each quarter	50% One fourth each quarter	75%	n.a.	66.67% One sixth each quarter	60% One fifth each quarter	50% One eighth each quarter	80% One fifth each quarter
<b>Data Collection</b>															
<i>% non-response Compuls./volunt.</i>	15% Voluntary	n.a. Compulsory	29% Voluntary	3% Voluntary	8-10% n.a	n.a. Voluntary	13,6% n.a.	7% Voluntary	5% Compuls.	22% Voluntary	40-50% Voluntary	9% Compuls.	15% Voluntary	12-14% Voluntary	22% (1 <sup>st</sup> . wave) 6% (2-5 waves) Voluntary
<b>Weighting procedures</b>															
<i>Variables of post stratification</i>	Sex, age, region and nationality	Sex, age and province	Gross inc., age and educ. (for people registered as unemployed) or industry (for non registered as unemployed)	Sex, age, region and nationality	None	Sex, age and region	Sex and age	Sex, age and region	Sex and age	Sex, age, nationality and size of household	Sex, age, region, nationality and marital status	Sex, age and region	Sex, age, region and labour force status	Sex, age and labour force status	Sex, age and region

Source: Martí and Ródenas (2007). For an actualization, see Eurostat (2007).