



9<sup>th</sup> Workshop on LFS methodology, Rome, 15-16 May 2014

## LFS data collection designs in Europe: present situation and proposal for the future

Arturo de la Fuente (Eurostat)

### 1. LFS DATA COLLECTION DESIGNS IN EUROPE: PRESENT SITUATION

Regulation 577/98 stipulates minimum requirements for the EU Labour Force Survey, e.g. as regards quarterly periodicity and continuous survey, existence of structural variables and ad-hoc modules. National LFS sample designs across European countries fulfil the EU minimum requirements but nevertheless show a lot of diversity. In that context, this paper focuses on the national solutions for collecting structural variables and ad-hoc modules.

As regards structural variables, some countries collect them in the whole sample whereas others collect them in a sub-sample. Normally this sub-sample is one or several waves.

As regards ad-hoc modules variables, some countries collect them in the whole sample, some in certain quarters and other countries collect them throughout the year but from part of the sample, normally using one or several waves. In some countries in this latter group, the part of the sample for the structural variables and ad-hoc modules is the same whereas in others it is not.

Countries approaches for sub-sampling to collect structural variables and ad-hoc modules will be shown with the following visual presentation: the sample for the whole year will be presented as a colour rectangle:



with the columns representing the quarters



Q1 Q2 Q3 Q4

and the rows representing the sample waves

wave 1	
wave 2	
wave ...	
wave n	

Chart 1 shows a simplified representation of the national practices regarding collection of structural variables and ad-hoc modules:

**Chart 1**

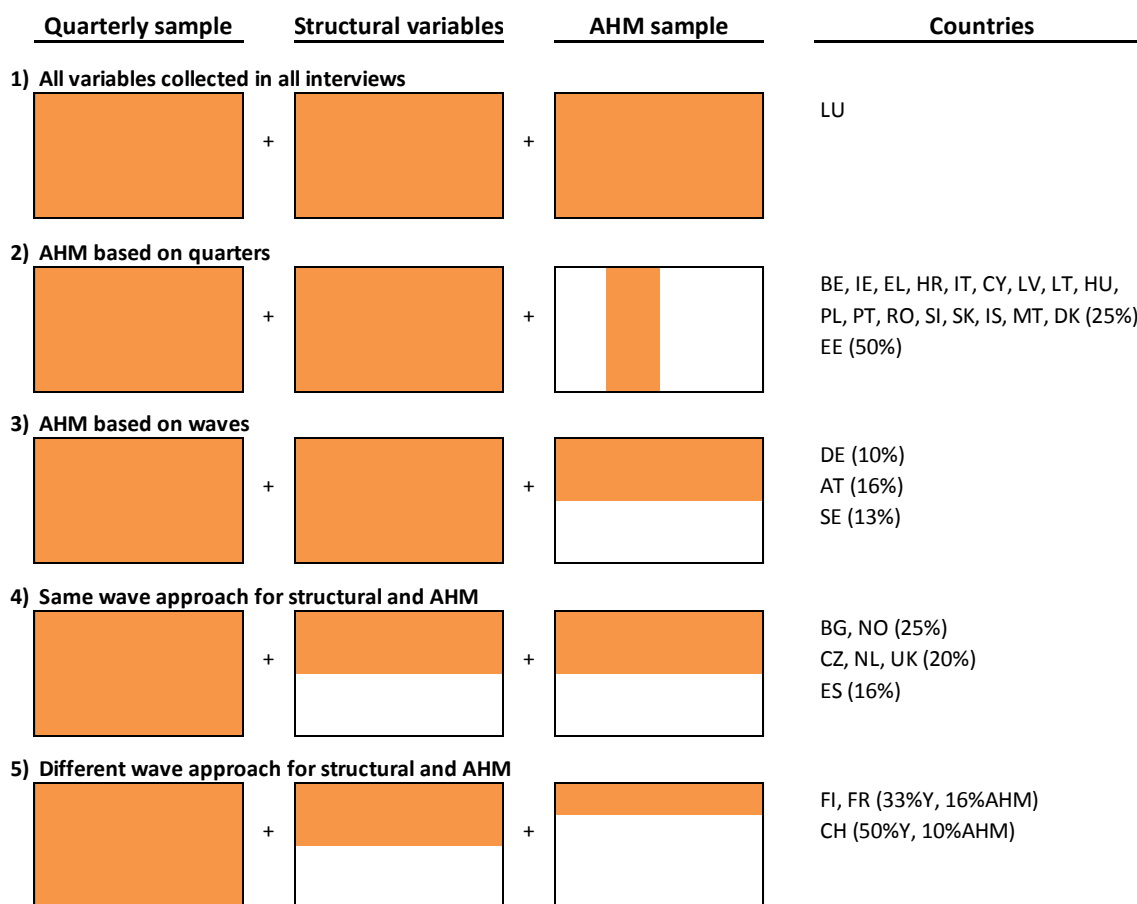


Chart 1 is detailed enough to give an idea of the national differences across Europe. However country practices differ even more. For instance chart 1 does not show the different number of waves in the countries:

<u>N. waves</u>	<u>Countries</u>	<u>Rotation pattern</u>
4 waves	BG,DK,EE,IT,LV,LT,MT,PL,RO, HR, CH	2-(2)-2
5 waves	CZ,IE,NL,AT,SK,UK	5-
5 waves	SI,FI	3-(1)-2
5 waves	IS	3-(2)-2
6 waves	GR,ES,FR,CY,HU,PT	6-
8 waves	SE,NO	8-

(DE, LU and BE do not presently have any infra-annual sample rotation scheme)

Chart 1 does not show either that the implementation of the wave approach for structural variables is not the same in all the countries in groups 4 and 5. In particular, some countries use it only for a few structural variables but not for all. Some Nordic countries do not collect the household variables (which are structural variables) in the same wave as the AHM. See the annex for these details.

## **2. ANALYSIS OF THE EXISTING SITUATION**

There are some lessons to be learnt from this situation:

- Only few countries collect structural (annual) variables from a sub-sample (so called wave approach). This means very few countries benefit from the advantages of the structural variables to limit response burden. It makes sense to develop a survey design for the LFS to facilitate collecting survey contents from only a part of the sample whenever it suffices the information needs. This is an efficient way of collecting data in a survey with a big sample like the LFS.
- The present implementation of the wave approach in some countries is too different from others, e.g. some countries applying it only for few variables (this is not visible in chart 1. See the annex). In order to fully exploit response burden reductions, the wave approach should be used for as many (ideally all) structural variables as possible. In addition, Eurostat can't handle different distinctions between non-structural and structural variables on individual country basis. This obliges Eurostat to suppress information in data extractions for fully surveyed structural variables. We need some general rules for a harmonised implementation of the wave approach at least in the data transmitted to Eurostat.
- The present implementation of the ad-hoc modules is too heterogeneous. Many countries do not collect data during the whole year, but in one quarter only. Whereas ad-hoc module variables typically do not have a seasonal pattern, results may show seasonality when ad-hoc module data are crossed with core LFS variables for analysis purposes.
- Only in few countries the collection of the ad-hoc modules is integrated with the structural variables. This integration would help to prevent errors. Presently the ad-hoc modules are very exposed to errors e.g. missed target groups which should have been included. This is a paradox because an 'ad-hoc module' is supposed to be easily mounted on the top of the (core) survey with limited burden and cost, i.e. just the opposite of what it actually is. It would be cheaper to mount them on the top of the structural variables than on the top of the quarterly variables.

The experience with the national sample designs for the structural variables and ad-hoc modules points to the need to give to the EU LFS a well-defined architecture for mounting contents to the survey to be collected only from a sub-sample. The modular architecture of the future LFS will provide a flexible framework to organise the survey contents. As the articulation of the LFS contents in modules is part of the agenda of the project modernisation of social statistics, we can take advantage of it for organizing the structural (annual) variables and the ad-hoc modules. That design must be flexible enough as to allow countries to use or easily adapt their existing practices (or to opt out from this architecture, i.e. asking all respondents all questions, but not to apply a different one). At the same time, it must be robust enough to be a joint framework used across all countries and serving the common needs.

## **3. A PROPOSAL FOR THE FUTURE**

Eurostat and the LAMAS working group have been working for some time on a possible new survey design for the EU LFS. This section reports some ideas from that discussion. This is still work in progress. The goal is to agree on the technical elements by the end of the year. Afterwards new EU legislation will be needed. Actual implementation in the LFS could be possible by the end of the decade.

The proposal is that the EU LFS carries quarterly modules, annual modules, biennial modules and rolling AHMs. The biennial modules are collected every second year. For the purpose here, they are considered annual 'placement modules' which are run or not depending on the year. The AHMs are also considered as annual 'placement modules' whose contents change every year.

The LFS modules are being created in terms of homogeneous content, same frequency and same target group. The aim is not that the modules have the same number of variables. Also for continuity reasons with the present LFS contents, the modules will be as long/short as necessary to accommodate variables. The biggest modules have some sub-modules. This sub-module structure facilitates possible implementation of those sub-modules in other social surveys.

- Table 1 below shows indicatively the possible modules for the future LFS. This is not a final proposal, it is work in progress. The intention is to focus on the modules as placeholders of contents here. Table 1 also includes the size of each module in terms of number of variables. This list consists of 21 modules, of which 11 are quarterly, 7 annual and 3 biennial plus a module containing the AHM called SAM (supplementary annual modules) in future.

This number is indicative as the modules are still in discussion.

**Table 1**

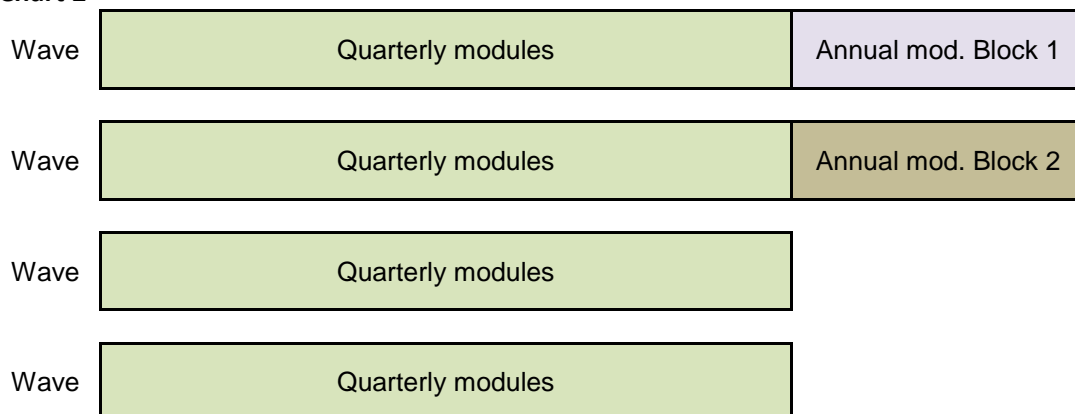
Modules and sub modules	Target group	Freq	Vars
1 Technical items	All		
2 Person characteristics	All	Q	7
3 Household characteristics	All	A	5
4 Socio economic background	Age>14	A	2
5 Employment status	Age>14	Q	4
6 Job characteristics	Main job	Q	7
7 Working hours	Main job	Q	5
8 Occupation	Main job	Q	2
9 Characteristics of the workplace	Main job	Q	5
10 Working time arrangements	Main job	2A	6
11 Start of job	Main job	Q	3
12 Second job	Working	A	4
13 Wished hours	Working	Q	5
14 Looking for other job	Working	A	3
15 Previous work experience	Not working	A	7
16 Looking for work	Not working	Q	11
17 Educational attainment	Age > 14	Q	5
18 Participation in education and training (4 weeks)	Age > 14	Q	2
19 Secondary information on participation	Age 15-34	A	3
20 Participation in education and training (12 months)	Age 18-69	2A	7
21 Health module	All	2A	2
22 Supplementary Annual Module	Variable	A	11

Eurostat proposed to LAMAS a survey design consisting in grouping the modules into blocks according to their periodicity. All the modules with quarterly periodicity would be in one block. The modules with annual, biennial and ad-hoc module variables would be placed in blocks separate from the quarterly block. In June and December 2013 Eurostat presented to LAMAS diverse options having 1, 2 or 3 blocks of annual modules, i.e. without overlaps between them.

What is the point of the blocks? The idea is associating blocks (i.e. survey contents) with waves (i.e. sample structure). This is the reason why the modules must be grouped. The block with quarterly modules is collected in all the waves i.e. from the whole LFS sample, whereas annual modules can be collected in only some waves (i.e. from a part of the sample only). The idea is to sub-sample the annual modules as a way to limit response burden. With this arrangement, annual modules, including AHM/SAMs, would be collected throughout the year but only in part of the sample.

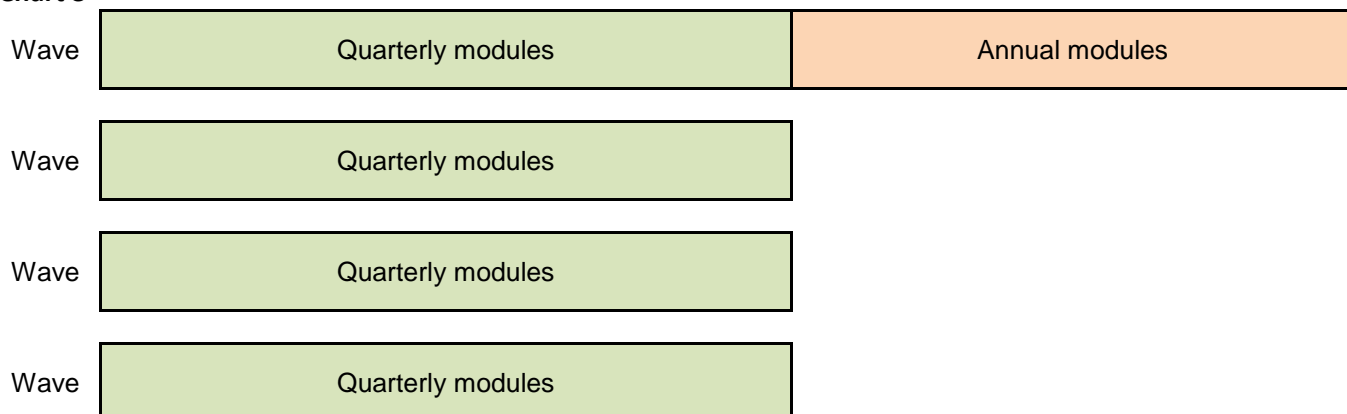
Chart 2 below shows schematically a framework with 2 blocks of annual modules:

**Chart 2**



For comparison, a framework with 1 single block of annual modules would look as follows:

**Chart 3**



(Charts 2 and 3 show a framework with 4 waves but the principle would be the same for surveys with a different number of waves. More about this below)

The main advantage of a system with 2 annual blocks, like shown in chart 2, is that the interview time will be shorter than if all the quarterly and annual variables were asked in the same interview. In a system with 1 annual block there would be one very long interview and the other repeated interviews would be shorter. In a system with 2 annual blocks the burden of the long interviews is spread in 2 interviews.

In countries not using the wave approach yet, this framework will increase the complexity of running the survey. This means more types of questionnaires, more complex logistics for questionnaire management, for IT for computer-assisted interviewing, more complex weighting and more types of weights. However some countries are already doing it that way. It is possible to share national experiences and to adopt solutions that can be easily used in other countries.

The allocation of annual modules in blocks is extremely important. In December 2013 LAMAS expressed preference for a system with 2 blocks of annual modules, like in chart 2. One possible allocation of modules in blocks would be as follows:

**Table 2**

Block 1 Additional background (BACKGR)	Block 2 Work details (JOBDET)
3. Household characteristics	10. Working time arrangements (biennial)
4. Socio-econ. background (i.e. mainstat + income)	12. Second jobs
15. Previous work experience	14. Looking for another job
Total: 14 variables	19. Secondary info on participation in education
	20. Participation in education and training in last 12 months
	21. health? (not decided yet)
	Total: 17 to 18 variables

The first block measures detailed background characteristics. It also includes the household composition variables. The second block measures detailed job characteristics not required on a quarterly basis. It includes the modules on working time arrangements (shift work, etc.), second job and looking for another job. This block would also include a few specific modules not directly related to the labour market that needs to be collected on a (multi-)annual basis.

In addition, each year the AHM/SAMs would be attached to one of those 2 blocks of modules depending on the content of the SAM that year. Examples of AHMs where the BACKGR block would be appropriate are: labour market situation of migrants (AHM 2014) and reconciliation of work and family life (AHM 2010). Examples of AHMs where the JOBDET block would be appropriate are: work organisation and working time arrangements (AHM 2015), self-employment (AHM 2017), accidents at work and work related health problems (AHM 2013) and young people on the labour market (AHM 2016). As result of this proposed design, AHMs would not be allocated to a specific quarter anymore, but would have to be collected all over the year. To which of the 2 annual blocks is attached the AHM/SAMs each year could be specified in the programme for AHMs.

One very important point for data exploitation in a system with 2 annual blocks is that the annual modules in different blocks cannot, in principle, be exploited together (but only with the quarterly variables collected in the same wave). Note however that all respondents will eventually pass through all the waves and be asked the 2 annual blocks. As a theoretical option, different annual blocks could still be used together with a microlinking technique, although they would refer to different moments in time.

A final, important issue is how the blocks (quarterly and annual) will be mounted in waves or groups of waves. Eurostat proposes to give a lot of flexibility to NSIs in this respect. Charts 2 and 3 above show 4 waves but they may correspond to groups of waves, to be decided nationally. Countries would be free to choose the order of the waves, as it fits to their national rotation scheme etc., provided they fulfil some precision requirements. Countries would also have flexibility to use different data collection modes for the waves, use dependent interviewing, etc. Countries are also free to accommodate their national variables and modules around this scheme.

Countries wishing to apply the two blocks in the same wave, i.e. as in chart 3 rather than chart 2, would of course be free to do it. That would be equivalent to having only 1 annual block.

If countries assign the BACKGR block to a different wave than the JOBDET block, the BACKGR block should be allocated to an earlier wave than the JOBDET sample for logical reasons. BACKGR is about background characteristics while JOBDET would aim to collect more detail, which more naturally fits with a later moment of data collection.

#### **4. CONCLUSION**

This paper explains a proposal of minimum requirements for national sample designs in a future modular LFS. This would consist of arranging the quarterly, annual and multiannual variables (including ad-hoc modules variables) into modules, and group the modules into blocks. There would be 2 blocks of annual and multiannual variables. These blocks determine the allocation of the annual/multiannual modules in the sample. It is essential that the same allocation of modules to blocks is followed by all countries that do sub-sampling, as to set conditions regarding which variables are collected together.

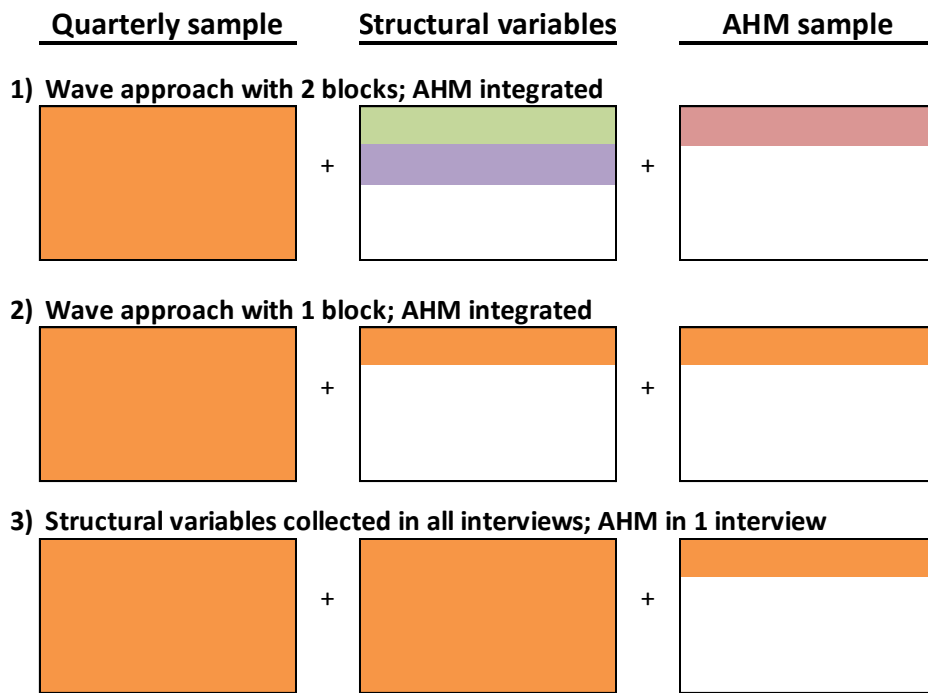
The blocks could be collected from one or more sub-samples covering the 52 reference weeks. It can be achieved by mounting the blocks in the survey waves, thus becoming an extension of the presently existing wave approach. This system would be compatible with multimode data collection systems.

The system would be quite rigid about which modules go in which blocks, but it would be rather flexible about which blocks go in which sample waves. Countries could keep their existing number of waves and infra-annual rotation patterns.

Chart 4 below presents visually this system (top part about wave approach with 2 blocks). As a fall-back position, countries wishing to do so could merge the 2 annual blocks, de facto working with 1 block. A final option not encouraged by Eurostat would be that countries not wishing to sub-sample the annual and multiannual modules could collect them in the whole sample, de facto opting out from the blocks system. This would require a separate solution to collect the AHM from a sample of independent observations, i.e. only once a year from each respondent, a solution not based on quarters.

This new system would overcome many of the flaws of the present situation and would take advantage of the future modular structure of the LFS.

Chart 4





## Annex: national implementations of wave approach

<b>Use of a sub-sample in 2012</b>	
<b>Country</b>	<b>List of yearly variables for which the wave approach is used (Ref.: Commission Reg. 377/2008, Annex II)</b>
<b>ES</b>	<p>The variables named below are collected yearly: SUPVISOR, SIZEFIRM, WAYFOUND, TEMPREAS, TEMPAGCY, SHIFTWK, EVENWK, NIGHTWK, SATWK, SUNWK, WAYMORE, HOMEWK, LOOKREAS, LEAVREAS, STAPROPR-NACEPR2D-ISCOPR3D (when the job finished more than 12 months ago), PRESEEK, EDUCFIELD, COURPURP, COURFILD, HATFIELD, WSTAT1Y, STAPRO1Y, NACE1Y2D, INCDECIL (from 2009 onwards), COEFFY.</p> <p>The variables named below are collected quarterly although they are sent yearly: HHLINK, HHSPOU, HHFATH, HHMOTH, MARSTAT, FTPTREAS, SEEKREAS, AVAIREAS, NEEDCARE, REGISTER, MAINSTAT, HATYEAR, COUNTR1Y, REGION1Y</p>
<b>FR</b>	only INCDECIL (c154161)
<b>NL</b>	<ul style="list-style-type: none"> <li>* All variables concerning atypical work (SHIFTWK, EVENWK, NIGHTWK, SATWK, SUNWK)</li> <li>* HOMEWK</li> <li>* All variables concerning previous work experience of person not in employment ( EXISTPR, YEARPR, MONTHPR, LEAVREAS, STAPROPR, NACEPR2D, ISOPR3D)</li> <li>* MAINSTAT</li> <li>* All variables concerning education or training successfully completed (HATLEVEL, HATFIELD, HATYEAR)</li> <li>* LFS ad hoc module</li> </ul>
<b>NO</b>	WAYJFOUN, FTPTREAS, TEMPREAS, TEMPAGCY, SHIFTWK, EVENWK, NIGHTWK, SATWK, SUNWK, WAYMORE, HOMEWK, LEAVREAS, SEEKREAS, AVAIREAS, PRESEEK, NEEDCARE, WSTAT1Y, STAPRO1Y, NACE1Y2D
<b>FI</b>	HHLINK, HHSPOU, HHFATH, HHMOTH, SUPVISOR, SIZEFIRM, WAYJFOUN, EVENWK, NIGHTWK, SATWK, SUNWK, WAYMORE, HOMEWK, LOOKREAS, SEEKREAS, AVAIREAS, PRESEEK, NEEDCARE, REGISTER, COURPURP, COURWORH, WSTAT1Y, STAPRO1Y, NACE1Y2D, COUNTR1Y, REGION1Y, INCDECIL, COEFFY, COEFFH, REG3D1Y. TEMPREAS has been changed to a quarterly variable.
<b>BG</b>	C36 (SUPVISOR), C37_38 (SIZEFIRM), C49 (WAYJFOUN), C51 (FTP TREAS), C53 (TEMPREAS), C56 (SHIFTWK), C57 (EVENWK), C58 (NIGHTWK), C59 (SATWK), C60 (SUNWK), C72 (WAYMORE), C75 (HOMEWK), C77 (LOOKREAS), C118 (AVAIREAS), C119 (PRESEEK), C120 (NEEDCARE), C121 (REGISTER), C122 (MAINSTAT), C125_127 (EDUCFIELD), C132 (COURPURP) C133_135 (COURFILD), C139_141 (HATFIELD), C142_C145 (HATYEAR), C141 (WSTAT1Y), C147 (STAPRO1Y), C148_149 (NACE1Y2D)
<b>UK</b>	No information available
<b>CH</b>	MARSTAT, SUPVISOR, SIZEFIRM, WAYJFOUN, FTPTREAS, TEMPREAS, TEMPAGCY, SHIFTWK, EVENWK, NIGHTWK, SATWK, SUNWK, WAYMORE, HOMEWK, LOOKREAS, LEAVREAS, SEEKREAS, AVAIREAS, PRESEEK, NEEDCARE, REGISTER, MAINSTAT, EDUCFIELD, COURPURP, HATFIELD, HATYEAR, WSTAT1Y, STAPRO1Y, NACE1Y2D, COUNTR1Y, REGION1Y, INCDECIL.
<b>CZ</b>	SUPVISOR, SIZEFIRM, WAYJFOUN, FTPTREAS, TEMPREAS, TEMPAGCY, SHIFTWK, EVENWK, NIGHTWK, SATWK, SUNWK, WAYMORE, HOMEWK, LOOKREAS, LEAVREAS, SEEKREAS, AVAIREAS, PRESEEK, NEEDCARE, REGISTER, HATFIELD, HATYEAR, WSTAT1Y, STAPRO1Y, NACE1Y2D_rev2, COUNTR1Y, REGION1Y, REG3D1Y