Administrative data and sample surveys’ data usage for determination of economic activity of population in register based Population and Housing Census in Latvia

Peteris Vegis, Deputy Director, Social Statistics Department, CSB of Latvia
Anna Klusa, Deputy Head of Population and Housing Census Section, CSB of Latvia

Geneva, Switzerland, 04 -06 October, 2017 Group of Experts on PHC19th meeting
Plan of the presentation

1. Introduction
2. Main administrative data sources that characterize the economic activity of the population in Latvia
3. Determination of the current activity status
4. Methodology of determining employment
5. Methodology of determining the main job
6. Occupation of employed persons
7. Industry / branch of economical activity (NACE)
8. Methodology on determination of status in employment
9. Currently economically not active population
10. Conclusions
Introduction

- 2021 Population and Housing Census in Latvia – register based Census
- Feasibility study on use of administrative data and information from statistical sample surveys to obtain data for the PHC 2021 section on economic characteristics of population:
  - current activity status
  - occupation
  - industry (branch of economic activity)
  - status in employment
- Administrative data storage in Social Statistics Data Warehouse
- A base for studies: Labour Force Survey (LFS) sample for 3rd and 4th quarter 2014 and 1st and 2nd quarter 2015. Comparison of estimations based on LFS data with corresponding information available from administrative data sources
Main administrative data sources that characterize the economic activity of the population in Latvia

Feasibility study included initial assessment of the availability and quality of administrative data in the databases of the:

- State Revenue Service (SRS)
- State Social Insurance Agency (SSIA)
- State Employment Agency (SEA)
- National Education Information System (NEIS)
- Data received from Higher Education Institutions
- Rural Support Service (RRS)
- Agricultural Data Centre (ADC)
- Database of legal and actual enterprises ([www.firmas.lv](http://www.firmas.lv))
- Business Register
Determination of the current activity status

- Latvian and international regulations analysed:
  - administrative data compilation process
  - definitions (e.g. national and international differences)
  - coverage of data

- Data sources identified

- Data from administrative sources compared with LFS data in aggregated level and individual level

- Work continued on base of administrative data on 1 January 2016
Economical activity of population in administrative data sources and LFS (January 1st 2016)

Geneva, Switzerland, 04-06 October, 2017

Group of Experts on PHC19th meeting
Methodology for determining employment based on administrative data

- In total data about 96% of employed persons found in administrative data sources. Some possible over-coverage should be evaluated
- Information from separate SRS data tables is combined by the PIN
- Residents working abroad: data collected from table on income obtained abroad by a natural person or by seafarer on a ship used for international transport
- SRS contains data of sufficient quality to count employed persons. Additional metadata necessary to complete timeliness, relevance and comparability assessment
- Statistical Farm Register gives contribution in identification of employed persons
Action plan of creating employment statistics

Select one status of economical activity per person

Select main job for each employed person (one taxpayer registration number)

Add occupation using personal id and taxpayer registration number
Add industry using taxpayer registration number
Add status in employment (in development)
Methodology for determining the main job

According to the information available in administrative data, the main job is:

• the one where the most hours have been worked.
• in cases when both the number of hours worked and the income were the same, the main job was determined randomly.
• the one with the largest income if the number of hours worked was not known for any of the jobs. If the income was identical, the main job was selected randomly.

Decision was made that persons who have salary tax booklet and submitted it for particular job have been working there most hours and meets the first criterion.
Distribution of employed persons by criterion/data source, %
(January 1st 2016)
Selecting the main job

Step 1. Get intermediate results from administrative data

22. Removed those who left job, calculation of maximal income
20. Quarterly income divided by three (3)
19. Detection of persons in maternity etc. leave
43. Removed those who left job, calculation of maximal income

26.1. Selection of income codes
26.2. Add indication that employed as sailor
27.1. Year income divided by 12
27.2. Year income divided by 12
28.2. Year income divided by 12
40. Selection of income codes, calculation of maximal income

FSS – selection of potentially employed

Step 2. Intermediate results are connected into one data set, adding to personal ID of inhabitants of Latvia on 01.01.2016. Main job is selected according to intermediate results about working hours, income and tax booklet.

Geneva, Switzerland, 04 -06 October, 2017
Main job:

✓ 78% of employed persons have just one job
✓ 18% - two jobs
✓ 4% - three and more
Occupation of employed persons

• SRS collects data about occupation since 2014
• SRS information of employees corresponds with the ILO methodology and covers 83% of all employed persons.
• Data not available for:
  • persons working in micro-enterprises
  • persons employed only on the basis of a work-performance contract
  • self-employed persons
• Compliance rate of occupation codes in SRS and LFS at the two-digit level is 61.5% and at the one-digit level – 68.9% (valued as satisfactory)
Occupation compliance in administrative data and LFS

• Occupations with the highest compliance between LFS and SRS were: dentists; physiotherapists; customs and border inspectors; locomotive engine drivers; bus and tram drivers; general medical practitioners etc.

• Occupations with the lowest compliance between LFS and SRS were: manufacturing supervisors; Web technicians; production clerks; handicraft workers not elsewhere classified; mixed crop growers etc.

• Occupations, where more precise information available in the SRS: police officers, accountants, teachers etc.
Missing data on occupations

Some data available from regular statistical surveys:
- LFS
- EU-SILC
- EHIS

Gathering data on occupation from institutions in 2017:
- Health Inspectorate (on medical practitioners)
- Notary Council
- Lawyer Council
- Bailiffs Council
- State Building Control Bureau

2020 Agriculture Census (possible adding of extra questions)
Industry / branch of economic activity (NACE)

- The NACE Section code was available from SBR or SRS data for 99.4% of employees that had the occupation code;
- The compliance rate between LFS and administrative data (on level of sections) accounted for 72.8% that is satisfactory;
- The highest level of compliance was observed in Section Q (Human health and social work activities) – 93.5%, in Section K (Financial and insurance activities) – 91.9%;
- The lowest compliance rate – 45.0% – was observed in Section N (Administrative and support service activities), followed by Section O (Public administration and defence; compulsory social security) – 45.9%.
- **NACE code 8411** (general public administration activities) should be studied more in detail.
- Methodology for imputation of missing values is worked out.
NACE compliance in administrative data and LFS (January 1\textsuperscript{st} 2016)
Statistical classification of economic activities (NACE)

On January 1\textsuperscript{st} 2016 the NACE code wasn’t detected for 12.1 thousands employees or 1.4\% (in January 1\textsuperscript{st} 2015 – 8.2\%)

\textbf{Problems:}

- Detecting the NACE in SRS data for persons employed in \textit{local government institutions/enterprises}
- Agriculture
- «Shadow economy»
Development of methodology on determination of status in employment

- The SIE is determined according to the information from the SRS taxpayers’ data and Business Register data on every inhabitant of Latvia. The indicators/variables are arranged in correspondence with the respective SRS data sources in the data file.

- Accurate data selection for determining the SIE in the main job was carried out only after it was determined whether each Latvian resident at working age is employed or not, as well as what is his or her main job.

- The process of combining data on persons main job and numbers of taxpayer registration codes of corresponding enterprises is done to detect whether the person is employer, employee or own-account worker.
Comparison of status in employment in administrative data and LFS
(January 1st 2015)

- Significant differences (24.7%) between LFS and administrative data among own-account workers:
  - 66% of them had the economic activity status «Other»
  - 20% were «Pension or capital income recipients»
  - 11% were «unemployed»
  - 4% - «not economically active students»

- Just 30% of employers were found in the administrative data in comparison with the estimate on the base of LFS

- Further research of specific cases (family doctors, lawyers, notaries etc.) is needed
### Status in employment

**(January 1, 2016)**

<table>
<thead>
<tr>
<th>Category</th>
<th>Administrative data</th>
<th>LFS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Count</td>
<td>%</td>
</tr>
<tr>
<td><strong>Employees</strong></td>
<td>752 690</td>
<td>86.2</td>
</tr>
<tr>
<td><strong>Employers</strong></td>
<td>34 029</td>
<td>3.9</td>
</tr>
<tr>
<td><strong>Own-account workers</strong></td>
<td>72 818</td>
<td>8.3</td>
</tr>
<tr>
<td><strong>Others</strong></td>
<td>13 454</td>
<td>1.5</td>
</tr>
<tr>
<td><strong>Total (+15)</strong></td>
<td>885 522</td>
<td>100</td>
</tr>
</tbody>
</table>
Methodology to determine the unemployment status

- Unemployed persons (99.8%) and job seekers (0.2%) registered by the SEA met the ILO definition of unemployed.
- Only persons registered with the SEA are included in the SSDW.
- SEA data represents typical age of unemployed persons (15 – 62) as just 1 - 2% of unemployed persons are older in accordance with the LFS.
- As LFS includes unregistered unemployment SEA data on 1st of January 2015 was by 21.5 thsd or 21.3% less than in the LFS estimate.
- In addition, part of the persons registered with the SEA as unemployed / job seaker in LFS were employed (4% of respondents) or inactive (23%).
Currently not economically active population

(Persons below working age (0-14); Pension recipients; Capital income recipients; Students; Others)

• Recipients of pensions from administrative data sources are set within the limits of the confidence interval and no significant changes in the method applied are required, including imputation
• Further study of income codes in SRS data is necessary to ensure more accurate selection of the data on capital income recipients
• Information on 21% of students was not acquired at individual level as the Register of Tertiary students in Latvia is not finished yet. Solution should be found
• Number of persons in the group «Others» will decrease with expected further improvements in other groups of not economically active population
Conclusions

• Employment in «shadow economy» should be studied
• Further studies of employees in agriculture are needed on base of administrative data and Agriculture Census 2020
• Identification of new data sources, analysis of their quality and follow-up to legislative changes in existing ones is very important
• The missing data imputation options should be investigated
• Forming an annual database of personal data on 1st January of the respective year (demographical data, economic activity, occupation, industry, education, etc.)
Conclusions

• Improvement of the Social Statistics Data Warehouse (SSDW) functionality should be continued to ensure the creation of Census data hypercubes

• The balance between the accuracy of statistics and the complexity of the system that produces statistics should be assessed

• Eurostat ESS VIP.ADMIN Grant project - a good base to obtain high quality economical characteristics of the population of Latvia from administrative data sources for the Census 2021
Thank you for your attention!

Geneva, Switzerland, 04-06 October, 2017