

Database for GPG analyse

Merle Paats
Statistics Estonia



Creating the database



Mapping available data sources

- Estonian Labour Force Survey (LFS) (ETU)
- Estonian EU-SILC (ESU)
- Census 2011 (REL)
- Estonian Tax register (EMTA)



- Personal and family characteristics and employment variables from surveys (ETU and ESU)
- Census: language skill and having the children (for women)
- Tax register: earnings from wage work

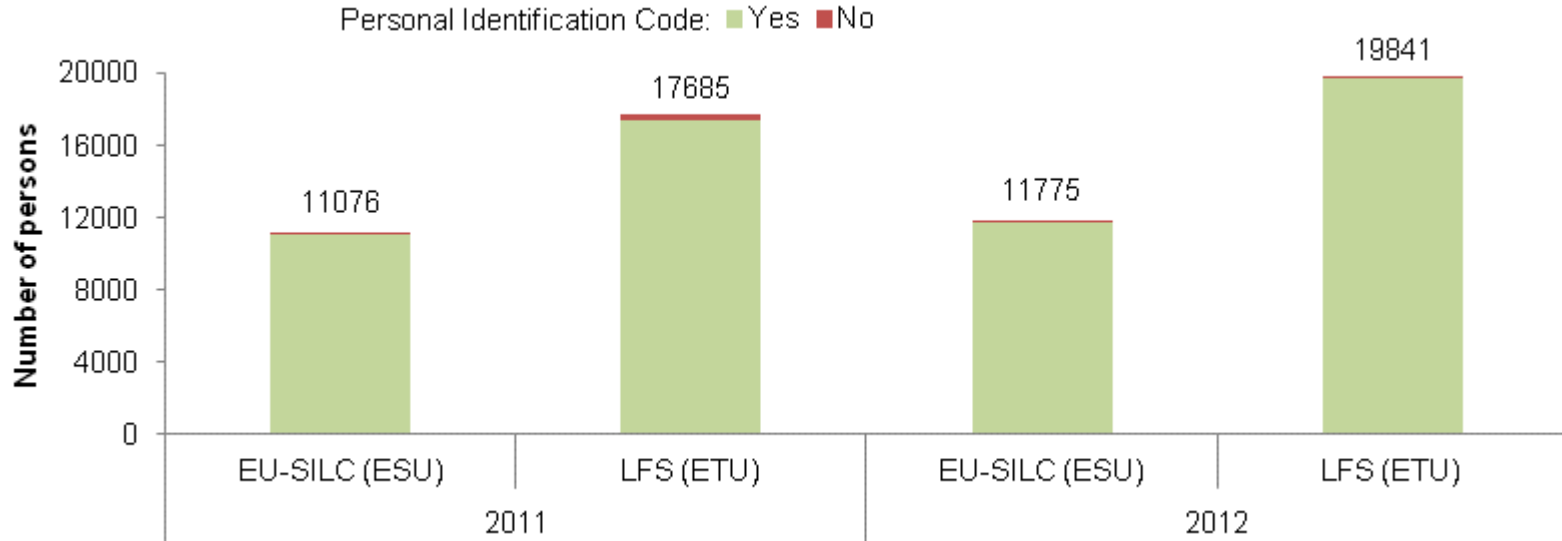
Actions for creating the database

- Compile the survey data:
 - Selecting the relevant variables
- Linking the data from:
 - Tax register
 - Census
- Calculating the weights (estimations on population level)

Problems: survey data

- Relevant variable existed in one survey only ⇒ Empty for other
- Relevant variable was collected differently in the surveys ⇒ Recoding
- Relevant variable was collected from different groups in the surveys ⇒ Filled only for employees

Problems in linking the survey and tax register data



Personal identification code is missing for less than 1% of respondents

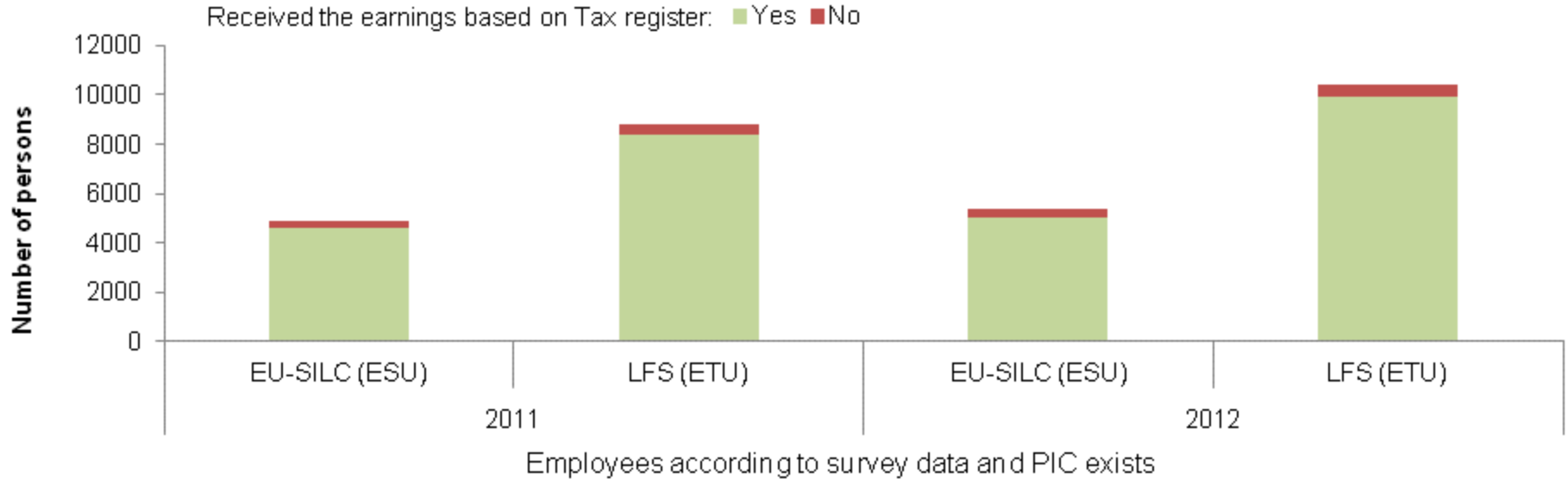
Problems in linking the survey and the Census

- Personal identification code is missing in the survey data
- Personal identification code is missing in the Census
- Different moment for conducting the survey and Census:
 - Persons participated in the survey in 2011 could be dead or left Estonia for the moment of Census
 - Persons participated in the survey in 2012 could be moved in Estonia after the Census or was younger than 15 in the moment of Census
- Data from Census 2011 are relevant for database 2011 and 2012, but not for earlier/later years

Main results



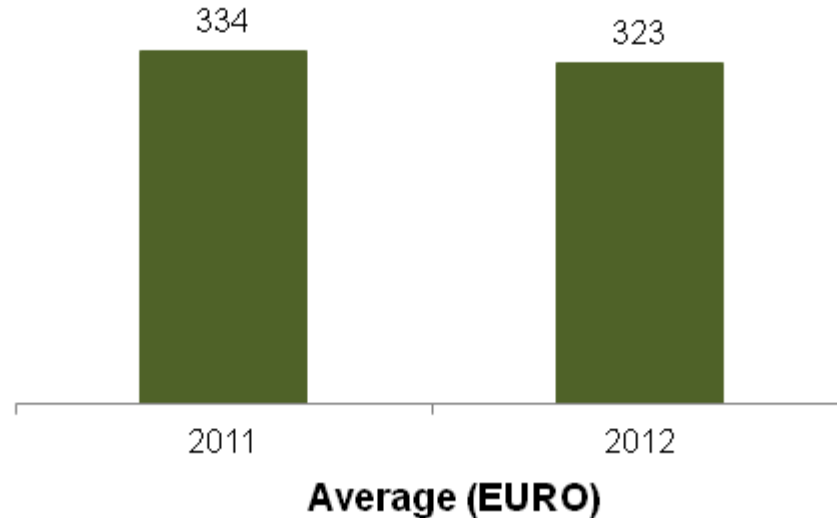
Data coverage in survey and tax register



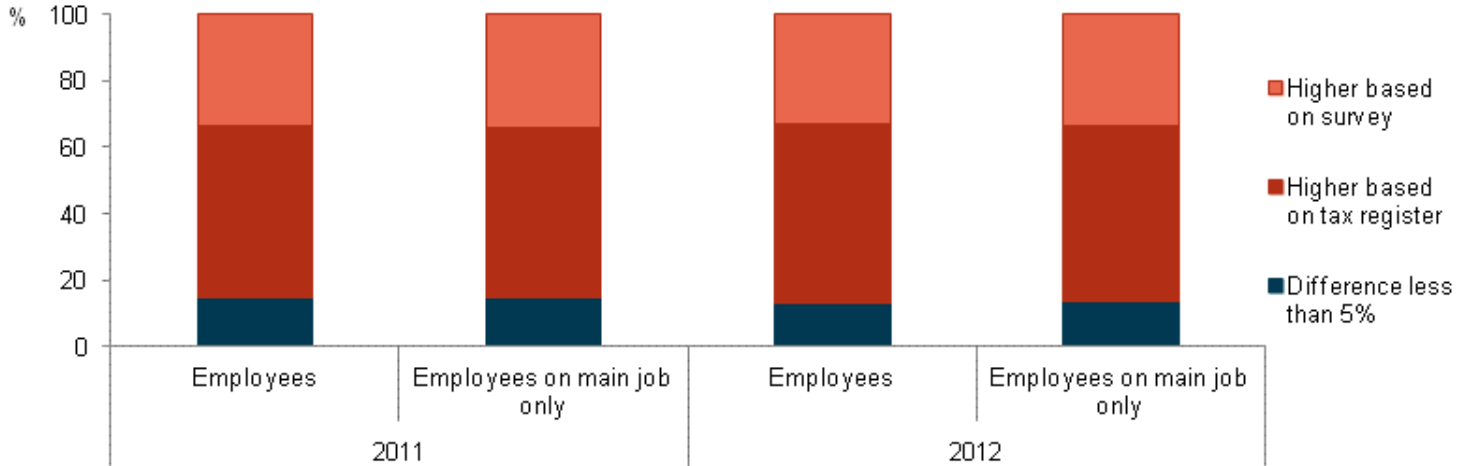
95% of employees (based on survey data) have received the earnings based on tax register

Data coverage in tax register and survey

- Less than 3% of respondents received earnings month after participating in the survey based on tax register, but said “not employed” in the survey
- Still, average earnings are smaller than usual and only one third of them received the earnings regularly during the survey year



Compatibility of earnings in survey and tax register



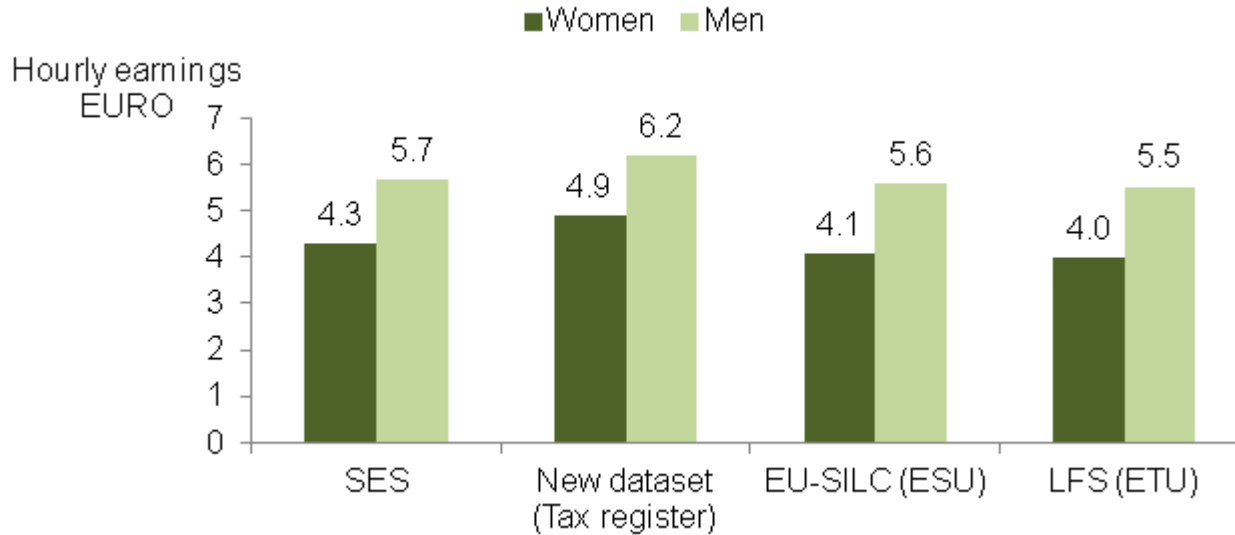
The income data based on survey and tax register are similar for less than 15% of persons who are employees based on survey and received the earnings based on tax register

Possible reasons: why data are different

- People do not want to tell real salary during the survey or do not show real situation formally (register data)
- Different reference period in the survey and register: reference week in survey vs yearly data in tax register
 - Earnings in the survey month
 - Earnings next month after the survey month
 - Average earnings during 3 months around the survey month
 - Average earnings during the survey year
- Different coverage: main employer in the survey vs several employers in tax register
 - Decide the main employer based on tax register data

Main conclusion: no difference

Initial results: hourly earnings, 2012



National GPG based on SES was 25%, based on new dataset: 21%, but is higher based on surveys: around 27%

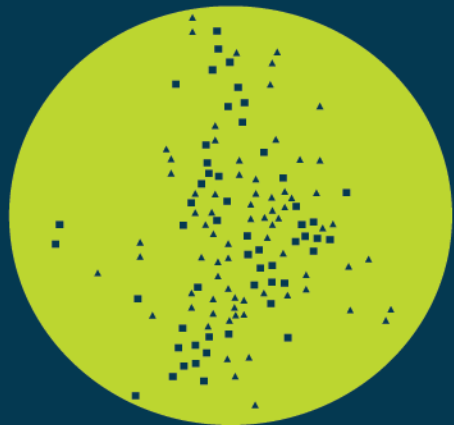
Possible reasons: why results are different

- Different reference period: data collected during 1 month (October) for SES 2012 vs average yearly earnings based on tax register
- Different definition for earnings in SES and new dataset
- Working hours: based on contract in SES vs usual working hours based on survey question

Main conclusion: more focus on working hours

Conclusions

- Database for years 2011 and 2012 were prepared for research done by Tartu University
- Based on lessons learned and main conclusions from TU study to prepare the databases for years 2006 – 2013
- Variable about working hours collected with surveys need to analyse more deeply and unrealistic values should be imputed



objekte kokku n 129
kolmnurki n 74
nelinurki n 54
ring n 1