17th Workshop on Labour Force Survey Methodology
Charlotta Olofsson, charlotta.olofsson@scb.se
Elin Ottosson Bixo, elin.ottossonbixo@scb.se
Statistics Sweden

Revised population in the Swedish LFS

Abstract

When implementing the new EU framework regulation on social statistics in the Swedish LFS from January 2021, several changes has been made to the survey. These changes regard the population, the definitions, and the questionnaire. The focus for this paper is the change in the population. The target population in the new framework regulation consists of people living in private households, as compared to the target population in the old framework regulation being the entire resident population. Due to this change, the target population decreased.

The estimation in the Swedish LFS builds on a generalized regression estimator (GREG) with auxiliary information from administrative data. The idea of using auxiliary information builds on the auxiliary variables co-varying with the survey variables and/or the response probability and/or that the auxiliary information identifies important study domains. The use of auxiliary information is intended to reduce sampling and nonresponse error. The base for the auxiliary information is Statistics Sweden's Total Population Register (TPR) that consists of those who are registered in Sweden. Before the new regulation, the TPR consisted of those who belonged to the target population. After the new regulation, the TPR does no longer comply with the target population of the LFS and the adjustment for the new target population consisting of private households was revised.

In Sweden, there is no prior information about the size of the population in private households. Therefore, the new target population was first achieved by estimates based on information collected in the interviews. As for all estimates based on a sample, these estimates were associated with an uncertainty. In this case, the estimates of the population became more volatile than what was the case before the new regulation. This was something that caused concern among the users of the statistics.

In order to reduce the volatility in the series of the population, work was carried out to find administrative data that could be used to identify individuals that do not live in private household. The aim with this work was to use this information in the estimation to adjust the auxiliary information so that it contains individuals in private households.

Three different groups were identified as making up the biggest part of those not living in private households. Administrative data, both on individual and aggregated level, on these groups were collected from both within and outside of Statistics Sweden. These administrative data was thereafter used to adjust the auxiliary information so that the total in the calibration only consists of individuals that live in private households. In this approach, the population no longer consists of an estimate but rather on administrative data. This leads to a reduction in the volatility in the series of the population. The new approach for the estimation was implemented in the end of 2023 and all estimates from January 2021 was adjusted according to the new approach.

SCB

Background

With the EU's new framework for social statistics being introduced in the LFS from January 2021, changes were implemented regarding definitions, questionnaires, and the target population the LFS. The target population in the LFS now consists of people aged 15-89 who live in private households. Compared to before, the change consists of an expansion of the age group from 15-74 years to 15-89 years and a delimitation based on type of household, which has not been previously occurred. From a Swedish perspective, people living in student corridors, lodgers, senior housing and serviced apartments is included, as it falls withing the scope of what Eurostat defines as private households. Those that are excluded from the target population are people who live in so-called institutional housing, such as group housing for the disabled, prisons and homes for closed youth care. Conscripts are also excluded from the target population.

Previously, the target population consisted of all persons registered in Sweden. These people can be identified through the Total Population Register (TPR). TPR has therefore, among other things, been used as a sample frame and as a basis for the creation of auxiliary information. The auxiliary information is used in the estimation procedure by adjusting the design weights so that the respondent's weights add up to known register totals. As there is no register information on the LFS's target population, those who do not live in private households have been mapped at the time of the interview, and then estimated from and including January 2021.

With this management, the population becomes an estimate, unlike before when it was based on register information, and hence it becomes more volatile. That the time series is more volatile when estimated is seen in Figure 1, which shows the population aged 15-74 for the period January 2018-August 2023. In the figure, the population for the period January 2018-December 2020 consists of all people registered, while that for the period January 2021-August 2023 consists of people registered living in private households. The figure shows that the population becomes lower according to the new target population and that it is more volatile compared to before.

Figure 1. Population 15−74 years according to the LFS.
Unlinked data, January 2018 - August 2023. Thousands.
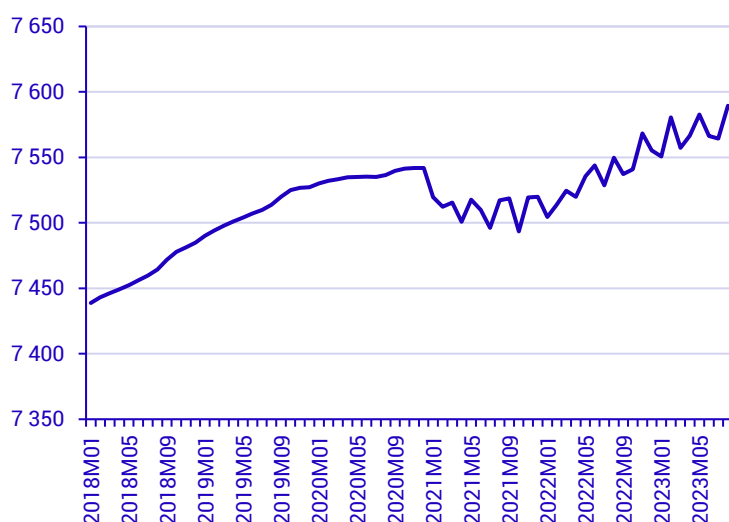
Source: Statistics Sweden, Labour Force Survey

Estimation

The estimation in the LFS builds on a generalized regression estimator (GREG) with auxiliary information from administrative data. The idea of using auxiliary information builds on the auxiliary variables co-varying with the survey variables and/or the response probability and/or that

the auxiliary information identifies important study domains. The use of auxiliary information is intended to reduce sampling and nonresponse error.

Regression estimation means that for sample individual k, an observation is made of (y_k, x_k) where y_k is an observed value from the survey while x_k is a vector of auxiliary information. The method also requires that the population total for the x-vector is known.

The auxiliary information in the LFS consists of register variables, or variables derived from register variables. Sources for the auxiliary information are the TPR and the monthly employer reports at individual level (AGI) as well as the register of the Swedish Public Employment Service. By using auxiliary information in the estimation, consistency is achieved between estimates in the LFS, and the known register totals used as auxiliary information.

New Procedure for the Population
In order to improve the quality of the LFS, work has been carried out to develop a new method for the LFS's target population. The focus of this work has been to use administrative data to create the population that forms the basis of the auxiliary information used in the estimation procedure. In connection to this work, parts of the LFS's help information have been revised for the purpose of quality improvement.

Target Population
Three groups have been identified as constituting the majority of those who do not belong to the LFS target group, these are:

 a) People who live in special housing, such as nursing homes, service homes or group housing
 b) Constript or contract soldiers
 c) Prison inmates

Special Housing
The Real Property Register contains address information for special housing. By co-processing this information with the population registration address in TPR, people in special housing can be identified. This information is available on a monthly basis with a lag of approximately one month when used in the LFS. Since the information is available on a monthly basis with high topicality, it is possible to adjust the population registered with regard to this information.

Conscripts
The Income and Taxation Register contains information about people who receive daily allowance for conscripts. These data are available annually with a lag of approximately one and a half years and are then divided by month. As data is made available with a long lag, it is not possible to use it to make adjustments at the individual level for the current month. Instead, the data for the month they actually refer to is used to calculate on an aggregated level how the various groups in the LFS's auxiliary information are affected. These aggregate data is then used to adjust the population for the current month.

Prison Inmates
The Swedish Prison and Probation Service has information on the number of ongoing prison sentences, and this corresponds to the number of inmates in prisons in Sweden. Statistics Sweden gets access to aggregated information from the Swedish Prison and Probation Service. This aggregated information refers to the first of the month and is delivered on a monthly basis. When used in the LFS, it is approximately two months old. The data is divided based on the auxiliary information used in the LFS and is therefore delivered by gender in combination with age and by regional division.

The auxiliary information in the LFS also includes information on country of birth. However, that information is not available at The Swedish Prison and Probation Service, and the delivery is therefore instead divided by country of citizenship. To convert the data by country of citizenship to country of birth, TPR has been used to model how these relate to each other and then recalculate country of citizenship to country of birth.

The Model

The three groups described in the previous section together constitute a large part of those who do not live in private households. These are therefore used to adjust the information from TPR regarding those who are registered in Sweden in order to identify the LFS's target population. This population model is equivalent to the population model used for microlinking of the years 2005–2020.

The data on special housing is used on an individual level for the specific month, while the data on conscripts and prison inmates is used on an aggregated level. These data are used to adjust the auxiliary information used in the LFS's estimation procedure.

The changed method means that the LFS's target population is now generated from register information instead of being estimated, which results in a less volatile time series over the population. This can be seen in Figure 2, which shows the population for the period January 2018 – August 2023. The period January 2018 – December 2020 consists of microlinked data and the period January 2021 – August 2023 consists of data produced according to the new estimation procedure. The figure shows that there is no longer a break between December 2020 and January 2021 because microlinked data is used before 2021. It is also seen that the time series from January 2021 is no longer as volatile as when it was produced through an estimation.

Figure 2. Population 15–74 years according to the LFS.
Linked data, January 2018 - August 2023. Thousands.
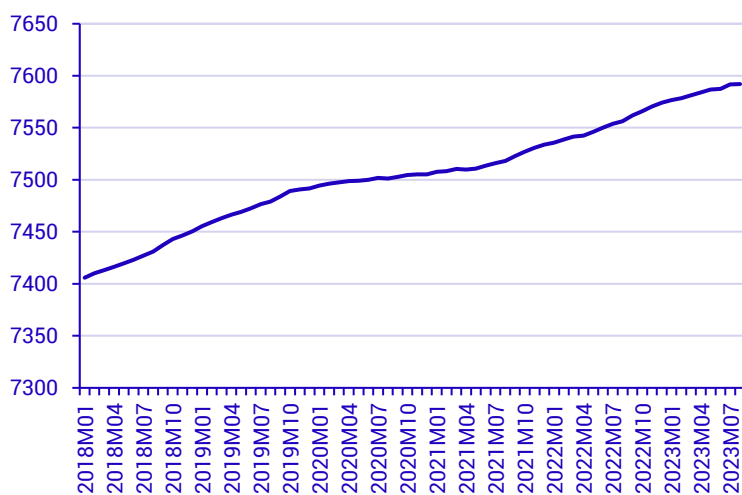
Source: Statistics Sweden, Labour Force Survey

Summary

The implementation of the new EU framework for social statistics brought on several changes compared with the old framework. One of the changes concern the definition of the target population that now consists of people aged 15–89 who live in private households. The change from resident population to private households means that the target population can no longer be found in administrative data. Therefore, the target population was estimated in the Swedish LFS when the new framework was implemented. This led to a more volatile series of the population than was the

case when the target population was the resident population. The reason for this is that the resident population can be found in administrative data and was therefore not estimated. A more volatile population was something that users raised concern about and although information on the total population in private households is not the main purpose of the LFS, a decrease in the quality for the population has an effect on the quality for the underlying groups. It is therefore of importance to maintain a high quality for the information on the target population.

To improve the quality of the information on the target population, work was carried out to identify administrative data that contains information on groups that do not belong to private households. Three groups were identified as constituting the majority of those who do not belong to private households. These three groups can be found in administrative data and constitute of people who live in special housings, conscript or contract soldiers and prison inmates. With the help of available administrative data on these three groups, a model was adopted in which these groups are excluded from the auxiliary information used in the estimation. By adopting this model where the auxiliary information is adjusted, the volatility in the population was reduced compared to when it was estimated.