

Re-developing cross-border commuting statistics for the Öresund region

Pernille Stender, Statistics Denmark, psd@dst.dk

Susanne Gullberg Brännström, Statistics Sweden, susanne.gullbergbrannstrom@scb.se

The Öresund Region is Northern Europe's largest labour market region with approx. 4 million inhabitants. After the opening of the Öresund Bridge in 2000, cross-border commuting in the region increased considerably. Therefore, there was a great demand for statistics that could shed light on commuting flows between the Danish and Swedish parts of the region. For this reason, Statistics Denmark and Statistics Sweden developed statistics on cross-border commuting which were released for the first time in 2000. The statistics (the previous cross-border statistics) were based on the exchange of data. The statistics published in 2000 held information about cross-border commuting in November 1998. Later, figures for 1997 were also published. Thereafter, the statistics were produced yearly until the release of figures for 2015 in 2017. Since then, it has not been possible to update the statistics. This is due to changes in administrative practice, which imply that it is no longer possible to exchange microdata at personal level between the two statistical agencies. This has been of great inconvenience to decision makers, analysts etc.

In the second half of 2021, Statistics Denmark and Statistics Sweden carried out a feasibility study. The study was funded by Region Skaane. The aim of the study was to investigate whether it would be possible to re-establish cross-border commuting statistics for the Öresund Region. The premise of the study was that it should be possible to produce the statistics without data exchange between the statistical offices. The results presented in this paper are preliminary and adjustments to the methods may still occur.

I. The previous method

The previous method was divided into three steps. They were:

First step: both agencies identified employees who were working in the country and who were not resident in the country. Data sources were the countries' register-based labour

market statistics (RAS and RAMS), where employment is recorded in November each year in both countries.

Second step: both agencies passed on the data to the other agency. The data passed on only had information about date, month and year of birth, sex, name and a serial number. Both agencies linked data (from the other agency) to their population register matching on date, month and year of birth, sex and name. When the person was found in the population register, each bureau enriched data with relevant background variables. These background variables were demographic variables, e.g. municipality of residence, education, place of birth and citizenship. When the person also had a job in the country of residence, information about salary in the job was added as well.

Third step: data was handed over to the other agency again. Information about date, month, year of birth and name was dropped. Additionally, variables related to the job and workplace (in the country where the person was not a resident), namely salary, industry, sector, municipality of place of work were added by the other agency. In case the person had a job in both Denmark and Sweden, the main job was identified. The main job was the job with the highest salary. A person was only counted as a commuter in the main job. Once both agencies had identified the cross-border commuters working in the country, the two statistical offices exchanged data again.

Quality issues in the previous statistics

The old method had some elements that contributed to uncertainty. In particular, the data processing in step two was a source of uncertainty, as in many cases the person's name could be spelled differently in the two countries. When Statistics Denmark for example performed the name match, the data was automatically divided into four groups according to how certain the name match was. In the first group, the name match was perfect. In the second group it was slightly uncertain, while in groups three and four it was uncertain. Manual checks were performed of groups two, three and four. In groups three and four, most records were typically rejected. In the second group, many records were accepted, but some were also rejected. For persons in group two, the middle name and surname were typically switched in some cases. Another typical occurrence was that the person had (perhaps) changed surname. In case the person had a more seldom first or middle name, it was

relatively certain that the person was the same. If, on the other hand, the person had a common first name, the decision became more uncertain and in reality very dependent on the decision-making rules of the person making the assessment.

Reason for discontinuation of the previous cross-border commuting statistics

As already mentioned, it has not been possible to exchange micro-data between the statistical agencies since 2017. This is due to changes in Swedish administrative practice, implying that Statistics Denmark assesses that there is a risk that Danish data could be disclosed for administrative decisions.

II. Re-development of cross-border statistics

Since autumn 2021, Statistics Denmark and Statistics Sweden have been working on developing new cross-border commuting statistics based on the premise that there will not be any exchange of data. The initial expectation was that statistics that are not based on data exchange would be of a poorer quality than the previous statistics.

Initially, as in the past, both Statistical bureaus used their register-based labour force statistics (RAS and RAMS) to identify employees working in the country and living abroad as a gross stock of cross-border commuters. Thereafter, an examination was made of national sources that could provide information on whether the person lived in the other country was done. Each bureau tried to develop the most appropriate method and the methods turned out not to be exactly the same.

The availability of the old cross-border commuting statistics has been a great advantage in the process for Statistics Denmark. For commuters from Sweden to Denmark, the previous cross-border statistics were used to make comparisons at an individual level between the methods. Initially, the hope was – of course - to get as close as possible to the figures in the previous statistics.

The Danish method

Since 2008, the e-Income Register has been the data source of the register-based labour force statistics (RAS). The e-Income Register has monthly information on employee jobs.

The e-Income Register has a very useful variable, which can be used to identify cross-border commuters. When an employee works in Denmark and lives abroad, the employer must report a country code for the country where the employee resides. This country code is the primary source for identifying cross-border commuters from Sweden to Denmark. Unfortunately, a number of employers do not report this information and therefore cross-border commuting statistics cannot be based solely on this variable. Furthermore, when there is no information about the country code, there is also no information on the address of residence. This implies that it is not known whether commuters live in the Swedish part of the Öresund Region.

The second source is information on citizenship from the population register. Persons with Swedish citizenship working in Denmark are also counted as cross-border commuters.

The third source is information on emigration from the population register. Persons who have emigrated from Denmark to Sweden and who have not re-emigrated are identified as cross-border commuters. This is an important source because many cross-border commuters working in Denmark and residing in Sweden are Danish citizens and therefore they cannot be identified by citizenship.

The majority of cross-border commuters to Denmark are identified on based on the country code. In 2020, 82 percent were identified by the country code (place of residence), 9 percent by citizenship and 9 percent by emigration from Denmark to Sweden. Of all commuters in 2020, 49 percent had emigrated from Denmark to Sweden and 49 percent held a Danish citizenship.

The Swedish method

Like the Danish method, Statistics Sweden uses data from the Swedish tax authority to identify the commuters. The data used is income declarations on individuals (AGI) which employers have reported monthly since 2019 instead of once a year, as was the case before 2019. From 2019, AGI is also used to produce the annual register based labour market

statistics (RAMS). Thereafter, different information is used to decide whether the employees live in Denmark.

The first source is information from the population register on citizenship and place of birth which is used to identify persons born in Denmark or with Danish citizenship who do not have a Swedish personal identification number but work in Sweden. These persons are given a special identification that can be linked to country of birth and citizenship. Persons who are born in Denmark or have Danish citizenship and do not live in Sweden are counted as cross-border commuters.

The second source is the country code. The country code is an important source for identifying cross-border commuters from Denmark to Sweden. Unfortunately, a number of employers do not report this information because it is only mandatory if the employer does not have information about the employees Swedish or special identification number. This means that cross-border commuting statistics cannot be based solely on the country code. Like in Denmark, when country code information is available, there will also be information on the address of residence, which makes it possible to deduce the municipality of residence. In this step, persons who are partly paying tax in a foreign country are added since this implies that they are living in Denmark.

At last the commuters identified by the first and second source are merged to remove duplicates. After the merge, approximately 25 percent of the commuters are identified only by citizenship/place of birth.

Differences between the Danish and Swedish methods

The methods chosen by the two statistical offices are not exactly the same. Statistics Sweden has chosen to use information on whether cross-border commuters from Denmark to Sweden have had foreign income in combination with the country code to determine the country of residence. Statistics Denmark relies on the country code itself. On the other hand, for Statistics Denmark, it has been valuable to use information about emigrations to Sweden, because many Danes have settled in the Swedish part of the Öresund Region and retained their work in the Danish part of the region.

Shortcomings of the new statistics compared to the previous one

The new cross-border commuting statistics have less complete background variables than the previous. However, a number of variables are equally complete: information on industry, sector, municipality of work, salary, and to some extent age, gender and citizenship.

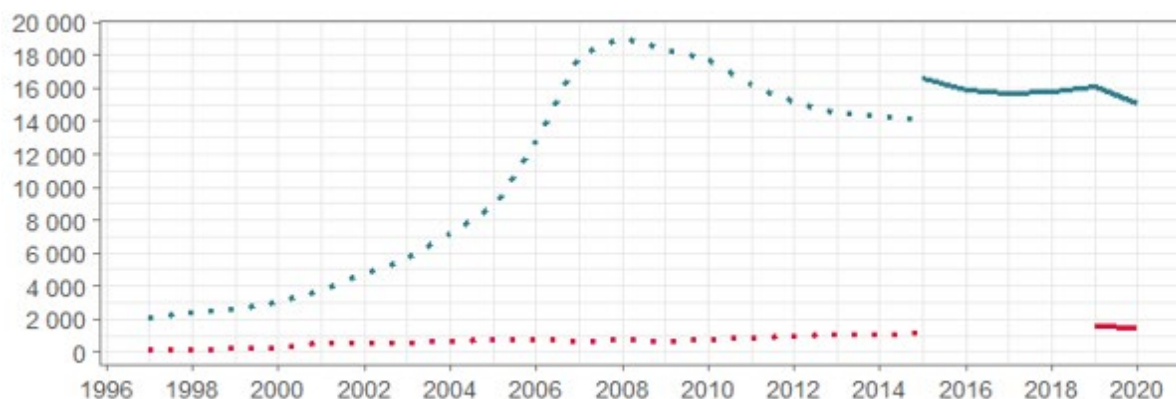
The new statistics do not have information on education because education is typically not completed in the country of work. Furthermore, the new statistics only have information on municipality of residence for commuters identified by country code. Therefore the statistics will probably be presented as cross-border commuting statistics from the whole of Sweden to the Danish part of the Öresund Region. The same limitation exists for the calculation of commuters from Denmark to Sweden.

In the new statistics, it is not possible to decide whether a person is a commuter in their main job in cases where they have jobs in both countries. Consequently, there will be that there will be cross-border commuters who in reality have their main job in the country of residence.

III. The results

The commuting flow from Sweden to Denmark is clearly the largest. The figure below is from a paper compiled by Region Skaane, which shows commuters from Denmark and Sweden in the Öresund Region for the period 1997–2020. The historical figures (the dashed graphs) are not fully comparable with the new figures. The old figures show commuters who both live and work in the Öresund Region, while the new figures show commuters who work in the region and live in the other country. This pulls the new figures in an upward direction.

Cross-border commuters in the Öresund Region



.. old definition, - new definition, - Sweden to Öresund, - Denmark to Öresund

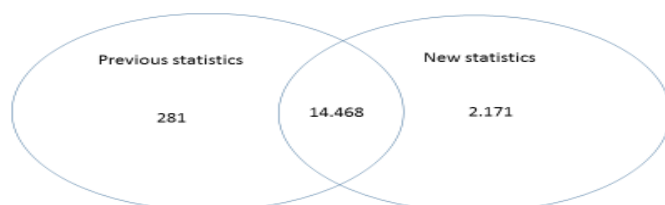
As can be seen, the figures for cross-border commuters from Denmark to Sweden are at roughly at the same level as in the previous Öresund statistics. In 2020, preliminary figures show that there were 2,400 commuters from Denmark to Sweden, and as shown in the figure, a little less than 2,000 of them works in the Öresund region. In 2020, there were 15,161 cross-border commuters from Sweden to the Danish part of the region.

Comparison between the new and previous statistics in 2015

The number of commuters from Sweden to Denmark is higher in 2015 in the new statistics. There were about 14,000 commuters from the Swedish to the Danish part of the region in 2015 in the previous statistics. Adding commuters living in the rest of Sweden, there were 14,749 commuters from all over Sweden to the Danish part of the region. In the new statistics the number of commuters is 16,639, which is equivalent to a difference of 1,890.

In order to ensure the quality of the new statistics, it has been useful to compare with the previous statistics at an individual level in 2015. The starting year of the new cross-border commuting statistics from Sweden to Denmark is 2015, because the country code is more complete in the e-Income register from 2015 onwards. At the same time, 2015 was the last year in which information was exchanged between Statistics Denmark and Statistics Sweden.

From Sweden to the Danish part of the region November 2015



The above figure shows the number of commuters in the previous and new statistics. The number of commuters is 1,890 higher in the new statistics. This covers the fact that the new statistics have 2,171 commuters who were not included in the old statistics. On the other hand, the old statistics had 281 commuters who are not included in the new statistics.

The result is encouraging for several reasons. Firstly, it shows that the criteria captures commuters quite well. The old statistics only had 281 commuters who could not be identified by country code, migration data or citizenship. The citizenship of 25 percent of these people was Polish, 15 percent were Danish, while 9 percent were British.

The new statistics, on the other hand, include 2,171 commuters who were not in the previous statistics. The question is whether too many commuters are identified. It has therefore been important to investigate the reasons for this difference. For 338 of the commuters, there is a good explanation. The reason is that the job in Denmark is their secondary job. For the remaining 1,844 persons, 1,325 (72 percent) of them had the country code SE, which is a strong indication that they are in fact cross-border commuters. For the remaining 519 persons, 289 had emigrated from Denmark to Sweden, which is also a strong indication of cross-border commuting. The remaining 230 (12 percent) were identified as commuters on the basis of citizenship.

Statistics Sweden has reassessed whether these persons in reality were resident in Sweden. That could be done without any further data exchange. Statistics Denmark sent their serial numbers in the previous statistics to Statistics Sweden. It turned out, that in reality the majority lived in Sweden. This pin points a weakness of the previous method, namely

the uncertainty of the name match. The problem was most likely the same in the matching process done in both Sweden and Denmark.

Statistics Sweden has not been able to conduct the same research as Statistics Denmark since AGI was first reported in 2019, and the information at hand could not be found in the previous yearly data from the Swedish Tax authority. There is no reason to believe that the conclusions on the Swedish part would differ severely from the Danish.

VI. Future perspectives

International perspectives

Several countries in Europe have experimented with compiling cross-border commuting statistics based on national sources only. An example is the pilot project “Border Region Data Collection” carried out by a consortium of statistical institutions, among them Statistics Denmark. In this project, the focus was very much on the use of administrative data sources. One of the conclusions was that “countries have different administrative systems with different sets of data and also that comparability of the data is not guaranteed”. The Nordic countries generally have very good administrative sources and for register-based employment statistics, they are – in many ways - very similar. This increases comparability and quality.

Sweden has borders to both Norway and Finland where commuting is frequent. The method used in Sweden can be used for border commuting statistics between Norway and Sweden and Sweden and Finland. The method used in Sweden can be applied to any country of origin for persons working in Sweden, although the most natural countries to start with would be the border countries since it is easier to define the country of residence in those cases.

However, inspired by the method, requested data from Eurostat on working persons from Ukraine will be compiled and delivered from Sweden using parts of the method. The difference is that we assume that these persons are refugees staying in Sweden and not resident in any other country.

The method used in Denmark can be used to compile cross-border commuting statistics from Germany to Denmark. Recently figures have been calculated for cross-border commuting from Germany to the Region of Southern Denmark. Cross-border commuters constitute a considerable part of the workforce in some municipalities in the Region of

Southern Denmark. Cross-border commuters are for example particularly important for the labour market in the municipality of Aabenraa. Here they increase the daytime population (people working in Aabenraa municipality) by 8 percent.

More timely statistics

So far, the new cross-border commuting statistics are based on the register-based structural statistics in Denmark, where the latest figures are for November 2020. However, the e-Income register has significantly more up-to-date information on the employed wage earners. For many years Statistics Denmark has compiled monthly and quarterly employment statistics based on e-Income, which could be used for more frequent cross-border commuting statistics. The population statistics also has very up-to-date information on population, emigration and citizenship.

In Sweden, the monthly data has been used during the whole year but the conditions concerning the company, like industry, refers to November. Sweden has developed new Register-based labour market statistics (BAS) launched in May 2022. Figures from BAS will be compiled quarterly in the start-up process and, can possibly be the base for more frequent commuting statistics.

Perhaps better data in the future

Cross-border commuters from Sweden to Denmark are primarily identified by the country code in the Danish e-Income Register. As mentioned earlier in some cases, employers do not report this code. Nevertheless, there is maybe hope ahead, because last year, the Danish Ministry of Taxation emphasised that employers should report country codes, addresses etc. for employees living abroad, in order to exchange income information with the person's country of residence.

V. Concluding remarks

We are confident, that a reliable cross-border commuting statistics have been re-developed. The plan is that at end of 2022, figures for commuting in the Öresund region will be published again in the Örestat databank. There is undoubtedly a great need for the data. Nevertheless, our hope is that it will be possible to exchange data again in the future. The aim would not

necessarily be to resume the production of cross-border commuting statistics on this basis, as we consider the quality of the newly developed statistics to be fairly high. However, there are some benefits when data exchange is possible, these are:

- better background variables (e.g. education)
- statistics that follows the concepts in labour market statistics (employment, main job)
- the possibility of integrating cross-border commuting statistics into national statistics, which was done with the previous Öresund statistics. This was highly requested by users

Data exchange would also make more complete ongoing checks possible, which could lead to adaptations of the methods. Data exchange would also improve knowledge sharing, which is – always - important.

References

Pernille Stender og Sara Ekmark (2015): *Pendlere og øvrige indkomstmodtagere i Øresundsregionen*

European commission (2018): *Border Region Data collection, Final report*