Border Region Data collection

Project n° 2016CE16BAT105

Final report

Written by Johan van der Valk (Statistics Netherlands)
September – 2018

Border Region Data collection

Project n° 2016CE16BAT105

Final Report

Prepared by: Johan van der Valk (Statistics Netherlands)

Anna Mitriaieva (Statistics Netherlands)
Wolfgang Seifert (IT.NRW)
Katharina Strauch (IT.NRW)
Mikael Skovbo (Statistics Denmark)
Ole Schnor (Statistics Denmark)
Marc Cierpial-Wolan (GUS)
Dorota Jasiukiewicz (GUS)
Igor Kuzma (Statistical Office of the Republic of Slovenia)
Samuel Balmand (INSEE)
Luisa Franconi (ISTAT)
with help of Maria Carla Congia, Elena Pontecorvo and Federica Pintaldi
Marc Debusschere (Statbel) assisted by Anja Termote

**Europe Direct is a service to help you find answers
to your questions about the European Union.**

Freephone number (*)**:**

00 800 6 7 8 9 10 11

(*)   The information given is free, as are most calls (though some operators, phone
        boxes or hotels may charge you).

Contents

Summary

This report is the result of a pilot project 'Cross-border data collection' carried out by a consortium of statistical institutes. The objective was to develop methodology to produce cross-border information. The consortium that carried out the project Cross-border Data Collection consists of the following statistical institutes: Statistics Belgium, INSEE (FR), IT.NRW (DE), ISTAT (IT), Statistical Office of the Republic of Slovenia (SI), GUS (PL), Statistics Denmark (DK) and Statistics Netherlands (NL) as project leader. The project took the theme of labour market as the focus of the attention as a case study to show what is possible. The project investigated the potential of three types of data sources in parallel: the Labour Force Survey (LFS), Administrative data and Mobile phone data.

The **Labour Force Survey** is an excellent source for harmonised data at national level but for border regions it does not allow the required geographical granularity to provide valuable evidence in the context of border regions. The limited sample sizes of the survey prevent the production of reliable NUTS 3 data for many regions in the larger countries. This is unfortunate because large countries have many borders. Alternatively, one can use NUTS 3 regions as building blocks to form larger border regions. As a matter of fact this is already done by Eurostat to provide data on EU border regions. In the project we have tested this approach and it seems a promising method delivering valuable information. We recommend to carry out further investigations to validate this method in the case of the LFS to prove its usefulness.

Administrative data is used extensively at national level, but what is possible at international level? All countries were able to provide data on incoming cross-border workers based social security files. Producing this data was not straightforward since the population of persons not resident in the country is a rather special population for which at national level less information is produced. Therefore additional work had to be carried out to provide this kind of information. We proved that it is possible and with relatively limited efforts. The results are very promising. We were able to provide information at NUTS 3 level on the share of employees that work in a specific regions and lives in the bordering country. This information enabled us together with data on citizenship to analyse cross-border work in the different border regions. With a simple set of indicators available for all border regions one can see how border regions differ. In many regions cross-border workers are of a 'classical' nature as border country citizens but in several regions a substantial part involves country nationals living across the border while having their job in their home country. We therefore recommend to put effort in organising the collection of this kind of data at a regular basis.

In addition, we recommend to complement the data that was collected with data from other countries and carry out a more comprehensive analysis of

the data. It what possible to do an in depth analysis of the data within the time frame of the project.

Mobile phone data has a lot of potential in the context of cross-border information. In this field measuring persons cross-borders is a key element to capture. Contrary to traditional data sources that generally measure static situations, mobile phone data captures movements. In addition, it allows for a standard methodology that can be used in principle worldwide. This would make it possible to produce international comparable information in an effective way. For that reason this kind of data is extremely powerful to produce cross-border information. We therefore recommend to give the production of cross-border information priority when developing statistics based on mobile phone data.

An analysis of Statistics Netherlands on Dutch mobile telephone data confirmed that the type data source has high potential. The movements of foreign mobile phones can be tracked into the country. It is in principle possible to distinguish cross-border workers, day-trippers and tourists from each other. This data source can provide consistent and timely information on all these important categories in the context of border regions. We are still at the beginning of the research regarding mobile phone data and much work has still to be done. Before we are able to produce official statistics with mobile phone data we still have to solve a number of issues. First of all, the methodology to produce the information has to be developed further. Secondly, to get access of the data is not straightforward. We recommend to organise a process coordinated and structural development of methods to exploit mobile phone data. NSI's can play an important role in this process to produce official statistics. We advise to take note of the work done in Work package 5 of the ESS-net project on Big data and follow up on the recommendation put forward in their report.

This project should be seen as a first step in a process to produce cross-border data on a structural basis. It is a proof of concept: a show case of what is possible. The data used and produced in the report does not have the status of a full-grown, mature product. A lot of work still has to be done to reach an acceptable level of completeness both in coverage of border regions and coverage of topics. Also the quality of the data and level of detail of the information presented in the report can and should be improved in a next phase of the process.

Apart from lessons learned on the technical side we have also learned about the organisational side. This project shows that a collaboration of statistical institutes is able to produce new information that is directly relevant for local, national and international policy in cost-effective way. It seems to be an approach that can be taken to develop more cross-border information. In addition to the development of statistics one has to think about how to implement the results. Developing a new product is one thing, putting it into production is another. We advise to make use of the current statistical

infrastructure of the European Statistical System and investigate how the production of cross-border data can be best integrated within this system.

1. Introduction

European cross-border cooperation is attracting more attention nowadays due to its influence on social cohesion, growth potential for border areas and nation member states, (labour) mobility of workers and knowledge exchange, and being one of the main drivers for overall harmonious EU development. Evidence of the increased policy interest is the EC communication from September 2017 'Boosting growth and cohesion in EU border regions'. More information is needed on how data collection for cross-border cooperation areas can be improved, with newly defined indicators for border areas and indices of (labour market) flows.

DG Regio of the E.C. decided to launch a pilot-project for statistical institutes to address this issue. This is mentioned under action point 10 on building evidence for better decision making. It should be seen as an example of a collaborative project on data collection initiatives on cross-border cooperation, in which indicators are captured in a systematic way for several countries at once, making cost-effective use of data sources (smaller footprint) that are already available, such as surveys and registers at National Statistics Institutes (NSIs) on for instance the labour market. Moreover and new methodological approaches for integrating those with new sources of data including so-called Big Data were tested.

The consortium that carried out the project Cross-border Data Collection consists of the following statistical institutes: Statistics Belgium, INSEE (FR), IT.NRW[1] (DE), ISTAT (IT), Statistical Office of the Republic of Slovenia (SI), GUS (PL), Statistics Denmark (DK) and Statistics Netherlands (NL) as project leader. The strategy was to develop a methodology for producing cross-border information in a cost-effective way making optimal use of existing data and expertise. The project took the theme of labour market as the focus of the attention as a case study to show what is possible. The project investigated the potential of three types of data sources in parallel: the Labour Force Survey (LFS), Administrative data and Mobile phone data.

This document is the final report for the project Cross-border data collection (CBDC). For all three lines of action we present the experiences in the process of data collection, results of analysis of the data and conclusions.

---

[1] IT.NRW is the statistical office of the federal state of North-Rhine Westphalia that provided the data for all German regions.

2. Labour Force Survey (LFS)

2.1. Introduction

LFS data is harmonised and available for all countries at EU-level. To what extent can we use this source to provide information in the context of cross-border work? In the project we try to answer this question.

2.2. Data collection process of LFS data

First we decided which LFS-data we would like to collect. The LFS should show how the structure of the labour market in the border regions is. One would like to compare this with the whole country and with other larger regions like NUTS2 or NUTS1 that are normally used for LFS statistics.

Regional scope
For the smaller countries we collected information on all NUTS 3 regions. This relates to BE, SI, DK and NL. For the larger countries (FR, DE, PL and IT) we collected information on all NUTS3 regions that are part of NUTS1 regions that are either (close to the) border of the countries considered. NUTS 3 regions are clustered into border regions and close to the border regions for all combinations of two countries having a border. We also took Luxemburg into account as country to be considered. We also collected Luxemburg data and made maps including this country. For this reason, we also distinguished (close to the) border regions with LU for BE, FR and DE.

Indicators
For the LFS indicators we took as reference the Eurostat Regional yearbook. This means that for unemployment the age group 15-74 years is considered and for employment rate 20-64 years. These age categories are used because they are Europe 2020 indicators with specific target. In order to be able to compare our results to the regional yearbook we think it is best to take the same age bands. For the age category 15-74 years we collected the following indicators: Population, Total labour force, Employed, Unemployed. For the specific groups of employed we use 20-64 years as age bands to be consistent with the standard EU indicators. That involves the following indicators: Total population 20-64 years, Employed 20-64 years, Self-employed, Employees, Employees with temporary contracts, Part-time workers.

Breakdowns
For NUTS3 regions we collected the indicators for the regions without breakdowns. For aggregates of regions we want breakdowns by sex, age and educational attainment. The following aggregates were included: NUTS0, NUTS1, NUTS2, border regions and close to the border regions. For the larger countries we collected data for NUTS1 and NUTS2 regions that are close to the border.

We translated the above description into a template per country that identify the border regions. That facilitated the data collection process a lot. The data

collection for LFS data went relatively smoothly. Data was delivered on time. For the project we used 2016 data but in principle 2017 data would be available in the course of 2018.

2.3. Results of the analysis of the LFS data

2.3.1. Data on NUTS 3 level

Most of the countries use thresholds for the LFS below which data cannot be published because of too few observations in the sample. For NUTS3 regions samples are generally quite small. If you then take as indicator a relative rare phenomenon like unemployment the chances are high the numbers are too small to publish.

Low number are a serious issue in the case of Germany with its low unemployment rate. We see that for Germany unemployment at NUTS 3 level cannot be published for many regions. The way the LFS is organised makes it even more problematic. The LFS is regulated by federal statistical office but the survey is conducted by the statistical offices of each Federal States. Sample sizes and data collection methods differ per state. Due to limitations of the sample size the federal states of two very important border regions, Rhineland-Palatinate and Baden-Württemberg, do not publish any data on NUTS3 level at all. For this reason, they did not provide data. This means that for Germany, country with the largest number of borders, the potential for describing the labour market at NUTS 3 level for Germany is very limited. The analysis INSEE carried out for using the LFS at NUTS3-level in France gave similar results as for Germany. They concluded that for a number of NUTS 3 regions sample sizes are too small to publish. They argue that only with combining the LFS with other sources reliable estimates can be produced at NUTS 3 level. Also for Poland the LFS cannot be used at NUTS 3 regional level.

Map 2.1 shows the availability of unemployment information for the countries The picture is quite disappointing. For many NUTS 3 regions sample sizes are too small to publish the data. Unemployment is in theory a good indicator of the situation of the local labour market. Therefore is it a pity that this data cannot be used at NUTS 3 level for many regions.

We can note that the sizes of the NUTS 3 regions differ quite strongly between the countries considered. For the German federal states North Rhine Westphalia, Rhineland-Palatinate, Saarland and Baden-Württemberg and Belgium these regions are much smaller than in France or Poland. This means that the granularity of NUTS 3 data differs considerably per region. And this in turn has consequences for the potential use of the LFS.

Map 2.1. Availability NUTS 3 LFS-data for unemployment in border regions
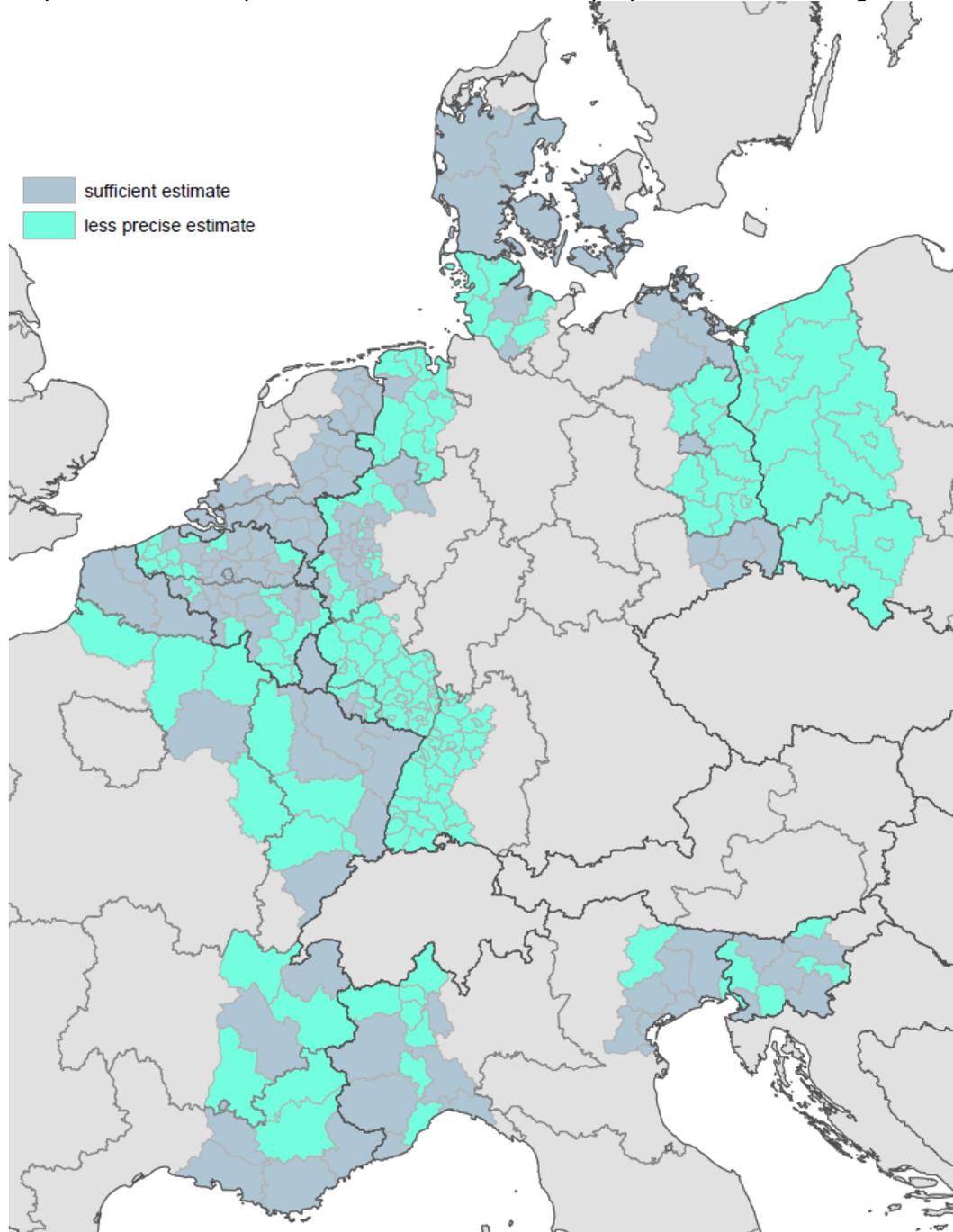
LFS can also be used to give information on more robust indicators like employment rates. Currently, this data is published by Eurostat at NUTS-2 level (See map 2.2).

Map 2.2 Employment rate 20-64 years at NUTS 2 level, LFS 2016

EU-28 = 71.1

	< 60
	60 – < 70
	70 – < 75
	75 – < 80
	>= 80
	Data not available
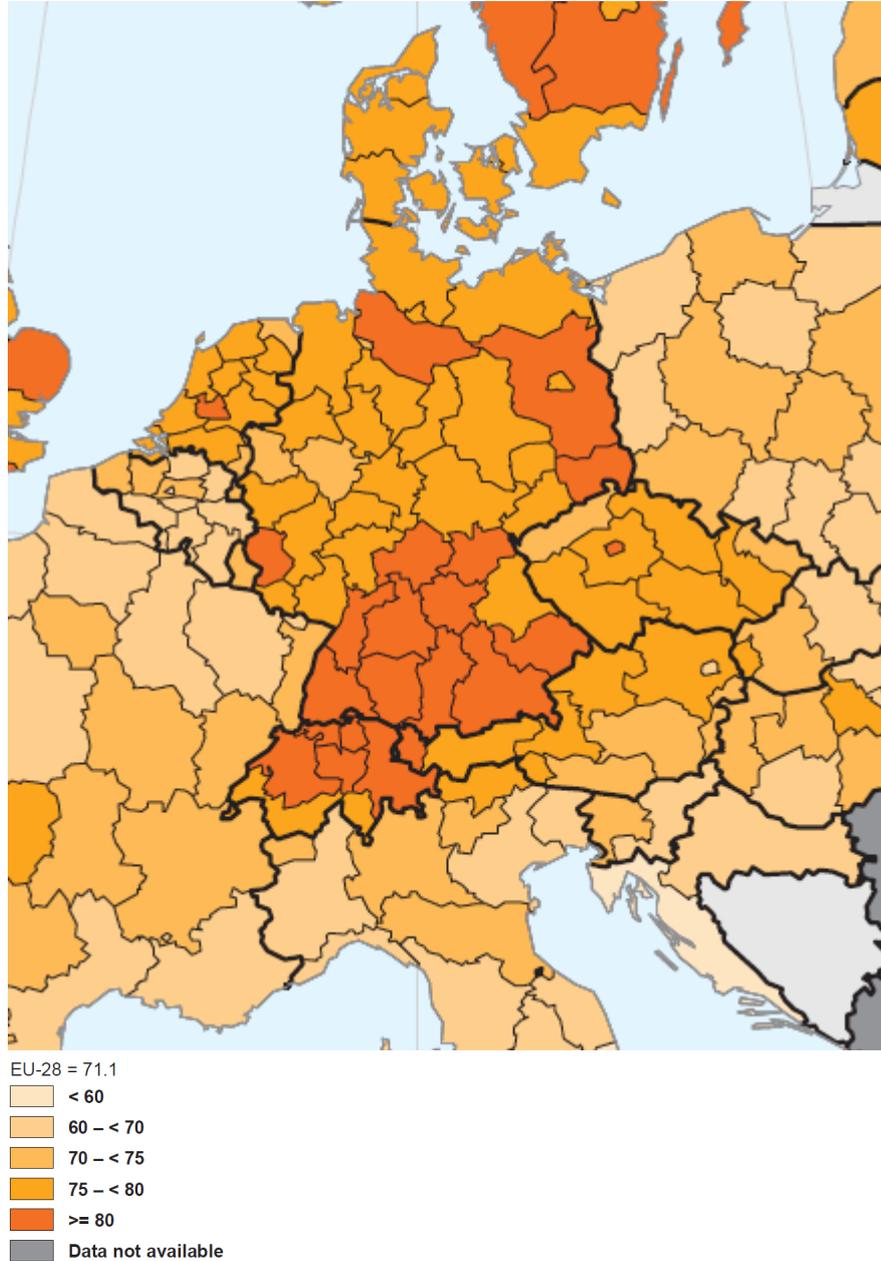
The interesting question is: what is the additional value of having this data available at NUTS 3 level? Map 2.3. presents the situation based on our data collection. It is unfortunate that data from Rhineland-Palatinate and Baden-Württemberg is missing. For the rest most border regions show relatively small differences in the employment rate. This suggests that data at NUTS 2 level is already quite useful and the added value of the NUTS 3 detail is limited.

Map 2.3. Employment rate 20-64 years at NUTS 3 level, LFS 2016
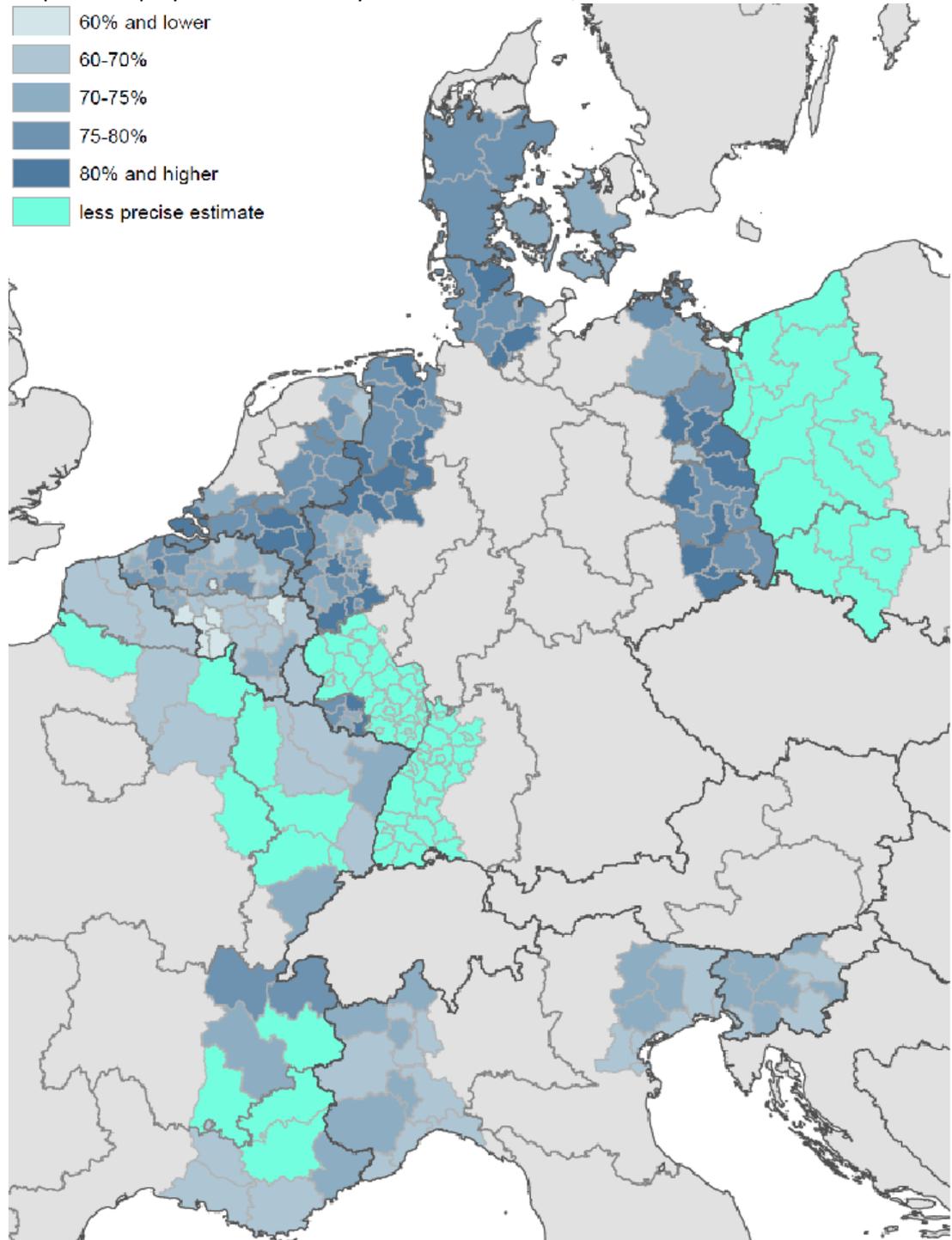
For part-time work the situation is similar. Map 2.4. shows that difference between NUTS 3 regions in a country is very small. Differences are prominent between countries. As a consequence NUTS 2 data is already quite valuable.

Map 2.4. Share of part-time work 20-64 years at NUTS 3 level, LFS 2016
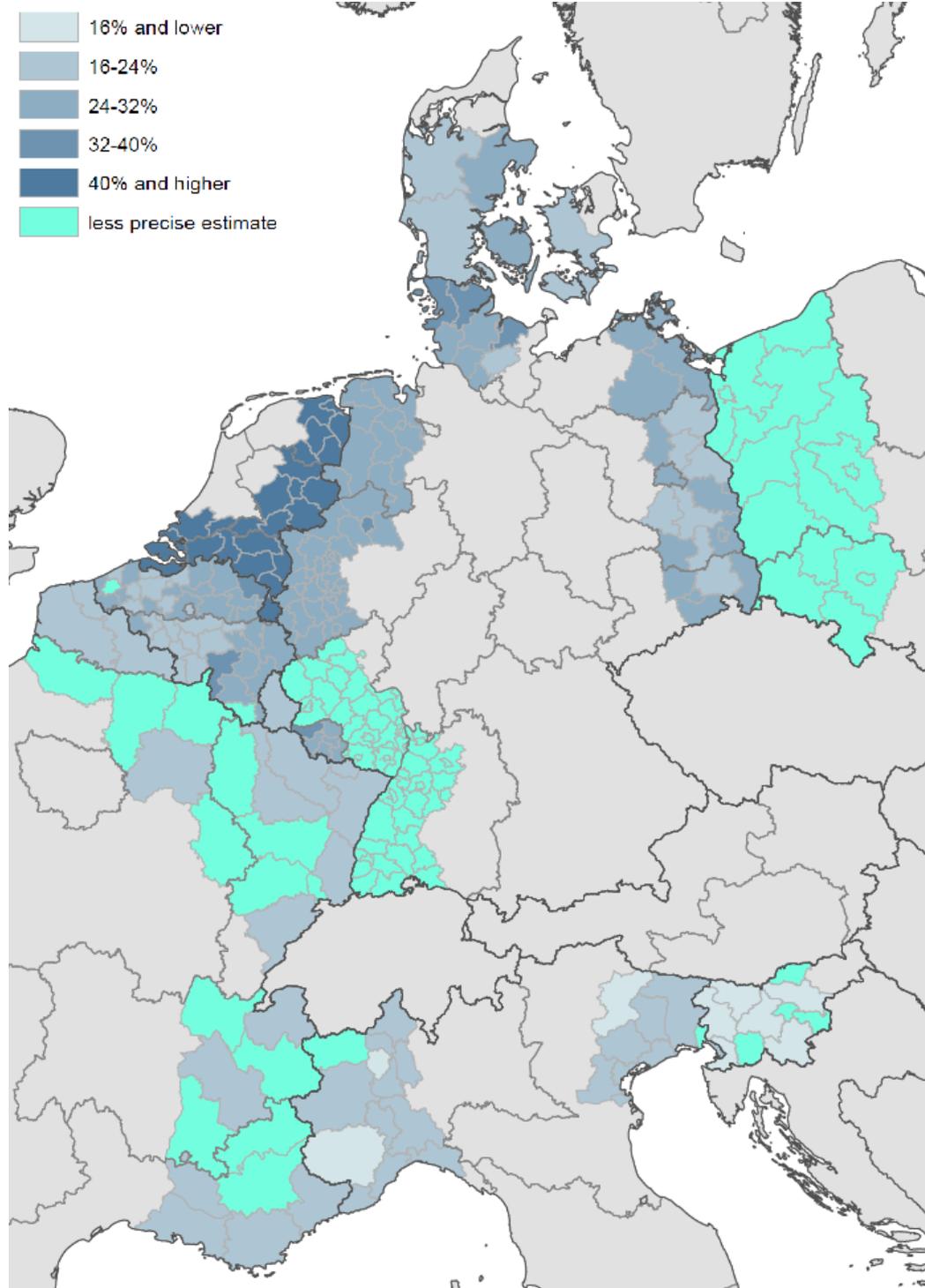
2.3.2. Data on border regions

Although NUTS 3 data is hardly usable to provide a picture on unemployment, they can be used as building blocks to say something on

border regions. Per border we have identified border regions and close-to-the-border regions as being the set of NUTS 3 regions at directly the border and the set of NUTS 3 regions adjacent to the border region being the second row of close-to-the-border regions. These regions are considerably larger than the original NUTS 3 regions. As a consequence the sample sizes are substantial and allow relatively reliable estimates based on the LFS. For instance for Germany and France the NUTS 3 LFS estimates on unemployment cannot be used in most cases but the border regions are large enough to provide sensible estimates. We also used these regions to allow for breakdowns by sex and age. In a similar way, aggregates could be computed for groups of border regions and close-to-the-border regions defined in an alternative way, i.e. including close-to-the-border regions that have at least 50% of their population living in an area of 25 km width along the border.

Figure 2.1 provides results for all border regions. The results show that in 2016 the unemployment rates are quite similar at both sides of the following borders: FR-IT, IT-SI, BE-NL and DE-PL. Some regions show some differences by sex and age. At the German side of the border unemployment rate are somewhat lower for women and persons below 45 years of age.

At the French side of the FR-BE border unemployment is considerably higher compared to the Belgian side. The FR-DE border also show a similar picture with relatively high rates at the French side. This would suggest that for French jobseekers in those regions the labour market at the other side of the border provides some potential. At the BE-DE border unemployment is much higher at the Belgian side of the border. In all those regions the differences are the largest for persons younger than 45 years of age.

In many cases the situation regarding unemployment differs not substantially between border regions on hand and close-to-the-border regions on the other hand. Belgium is the main exception where the unemployment in Walloon is much higher than in Flanders which is reflected in the (close to the) border regions with DE and NL.

Figure 2.1. Unemployment rate for border regions, LFS 2016

Unemployment rate border region FR-BE

Unemployment rate border region FR-IT

Unemployment rate border region FR-DE

Unemployment rate border regions IT-SI

Unemployment rate border region DE-DK

DK Close to the border · DK Border · DE Border · DE Close to the border

Unemployment rate border region DE-NL

NL Close to the border · NL Border · DE Border · DE Close to the border

Unemployment rate border region BE-NL

BE Close to the border · BE Border · NL Border · NL Close to the border

Unemployment rate border region BE-DE

BE Close to the border · BE Border · DE Border · DE Close to the border

Unemployment rate border regions DE-PL

2.4. Conclusions regarding the LFS

For cross-border data one would ideally like to have data on NUTS 3 level or preferably even on a more detailed geographical level. The LFS is the largest EU survey on labour market information. We investigated to what extent this source can be used to produce indicators at NUTS 3 level. The results are not very convincing. Unemployment is an extremely important indicator to assess the labour market situation. The LFS is the appropriate source to measure unemployment in an international harmonised way. Unfortunately, for several German federal states, for many regions in France and in Poland the sample sizes are too small to provide reliable estimates for NUTS 3 regions. So for too many regions this data is not directly usable.

Using NUTS 3 regions as building blocks to provide data on border regions (border regions and close-to-the-border regions for instance or the EU border regions as defined by DG Regio) could be a sensible alternative approach to make the best of the LFS data. Such regions have are substantially large corresponding with significant sample sizes for the LFS, that could be large enough to produce reliable estimates. It may allow estimation of unemployment rates and breakdowns like sex and age. For the participating countries this provided interesting results. It revealed the borders where the labour market situation on one side of the border differs strongly from the situation at the other side of the border. This was the case for the North and East borders of France and the South and East part of Belgium where the unemployment is much higher compared to the other side. We can conclude that using NUTS 3 regions as building blocks is a promising approach. Some methodological issues still have to be discussed before a final recommendation can be made. Since samples and weighting schemes were not designed to produced estimates for these constructed regions we have to investigate to what extent the results are representative. Given the implications on the operations that rely on the LFS adapting the design of the survey to accommodate for this is not opportune from our point of view.

An important application of using NUTS 3 regions as building blocks is to use it to produce information border regions as defined by DG regio of the E.C. These are defined as regions that have at least 50% of their population covered by the 25 km border areas. More information is available [here](). Eurostat is producing data for these kinds of regions.

Although unemployment rates cannot be used for many NUTS 3 regions, more robust indicators like the employment rate can be produced for those regions. However for these indicators the added value of NUTS 3 information compared to the available Eurostat data on NUTS 2 level is rather limited. For these kind of indicators he variation between NUTS 3 regions is namely small. Approximating such indicators on the basis of NUTS 2 information would be an cost-effective alternative of reasonable quality.

Finally, we would like to note that LFS data can be used as input for small area estimation. In countries like France, the Netherlands and Poland this a quite common approach. They combine LFS information with administrative data to provide tables on unemployment by NUTS 3 region. Annex I presents some work that was carried out by the statistical office of Poland (GUS) in this area.

3. Administrative data

3.1. Introduction

The second line of activity within this project was to investigate the collection of information on cross-border workers from administrative sources. This would in theory allow deeper regional detail compared with LFS data. In the project we collected data and identified the issues that occurred in this process. In addition, we analysed the results in order to see what the value is of the effort to collect this information.

3.2. Data collection process administrative data

3.2.1 Defining the data to be collected

At first we decided which administrative data we could collect within the scope of the project. A consultation on the availability of data showed that all countries in the consortium could provide administrative data on the population by citizenship and incoming cross-border workers and by country of work. This data would provide information on the structure of the population in the border regions and allows identifying cross-border workers. We decide to collect two data sets, data set A on the resident population and data set B on x-border workers.

Regional scope
For the smaller countries we collected information on all NUTS 3 regions. This relates to BE, SI, DK and NL. For the larger countries (FR, DE, PL and IT) we collected information on all NUTS3 regions that are part of NUTS1 regions that are either (close to the) border of the countries considered. We complemented this set with LU data as far as available on line.

Breakdowns
We included breakdowns by sex and citizenship. For data set A describing the resident border population we included age as well. For country of residence and citizenship we asked to identify all bordering countries. (see table 3.1 below).

Table 3.1

Country											
	(Border) countries to identify										
DE	DE	NL	BE	LU	FR	CH	AT	CZ	PL	DK	Other
NL	NL	DE	BE	Other							
BE	BE	FR	DE	LU	Other						
FR	FR	BE	LU	DE	CH	IT	ES	Other			
IT	IT	FR	CH	AT	SI	Other					
SI	SI	IT	AT	HU	HR	Other					
PL	PL	DE	CZ	SK	LT	BY	UA	RU	Other		
DK	DK	SE	DE	Other							

To identify cross-border workers we have to know where persons live and where they work. For this we need to know the place of residence and place of work. This is required for both the country at hand and the neighbouring country. Origin can be measured

using either country of birth or citizenship. We used the latter. Theoretically one would like to identify cross-border workers as presented in the table 3.2 below.

Table 3.2. Ideal information to identify cross-border workers

			Working				Country B		Other		Not-working
			Country A				Country B		Other		
			Border region	Other regions		Border region	Other regions				
Living											
Country A											
	Border region										
		Citizenship A				Outgoing					
		Citizenship B				x-border					
		Other				workers					
	Other regions										
		Citizenship A									
		Citizenship B									
		Other									
Country B											
	Border region										
		Citizenship A	Incoming								
		Citizenship B	x-border								
		Other	workers								
	Other regions										
		Citizenship A									
		Citizenship B									
		Other									
Other											
		Citizenship A									
		Citizenship B									
		Other									

In practice, to have this data for all countries proved to be too ambitious. In the kick-off meeting it became clear that administrative data on incoming cross-border workers is generally available in administrative sources. The required information is generally available in different administrative sources like population registers and social security files. Self-employed are difficult to capture through administrative sources. Therefore for pragmatic reasons we limited ourselves to employees to identify cross-border workers. In addition, the exact place of residence is unknown for incoming cross-border workers for virtually all countries. Therefore we settled for the country of residence only. The information that we collected to approximate cross-border workers can be presented in the following table 3.3 which is a simplified version of the theoretically ideal picture of table 3.2.

Table 3.3. Data on collected to approximate the identification of cross-border workers

			Employees working in	
			Country A	
			Border region	Other regions
Living				
Country A				
	Border region			
		Citizenship A		
		Citizenship B		
		Other		
	Other regions			
		Citizenship A		
		Citizenship B		
		Other		
Country B				
		Citizenship A	Incoming	
		Citizenship B	x-border	
		Other	workers	
Other				
		Citizenship A		
		Citizenship B		
		Other		

In addition to the data set on cross-border workers we also collected data on border population. We would like to see how many border country citizens live in border regions. We wanted to see if this data is available in the participating countries, what the quality of this data is and what the added value is compared to existing data and the social security data on cross-border workers.

3.2.1. Collecting the data

Contrary to the LFS data collection, collecting administrative data was not straightforward at all. Data arrived late, later or in one case was even not provided at all. Furthermore, data was in several cases incomplete or with specific issues. And finally data was not in all cases supplied according to the template that was agreed upon.

Table 3.4 shows a summary of the most important issues that was relevant to explain limitations in the results. For France, citizenship was not available. Belgium did not produce data on cross-borders due to limitations of national procedures to acquire administrative files. Statbel has no access to social security files. This data has to be applied for involving a time consuming procedure that prevented Statbel to deliver data within the time line of the project. Incoming workers residing outside of Belgium obviously are not in the population register, so the person's characteristics (specifically citizenship/nationality and place of residence) have to be obtained elsewhere. They are in the so-called 'BIS register' to which Statistics Belgium did not yet have access. The procedure to obtain access, unfortunately, is extremely complicated and time-

consuming. It used to involve an authorisation by the Belgian Privacy Commission, the procurement of which might take 6 to 12 months. To make things worse, the GDPR entry in force has cancelled this procedure and at this moment it is not clear yet what the new procedure might be. As a result, Statbel found it impossible to get hold of the crucial variables although these data do exist and should be obtainable in the future. An additional problem in this case, however, might be confidentiality: relatively small numbers to begin with (incoming workers by NUTS-3 workplace and place of residence) are further broken down by sex and nationality, resulting in numerous cells with very small numbers which cannot be communicated without additional processing. For the analysis we used data for Belgium that is available as open data. This had as limitation that not all NUTS3 regions were covered correctly and that citizenship was missing.

The good news is that in many cases data for 2016 was available or could be made available. In the context of this project we accepted data on cross-border data from 2015 just to be sure to have the data in time. But in principle data can be produced with a timing of a bit more than one year after the reference year.

Table 3.4. Summary of data collection issues for administrative data

X-border workers		
	Reference year	Issues
NL	2015	
BE	2015	Data set not produced, no citizenship available
DE	2016	Civil cervants and persons with a minor job are excluded
DK	2015	
FR	2015	No citizenship available and country of residence required to combine data from serveral sources
IT	2015	No specfication of country of residence is available
PL	2016	
SI	2016	

Border Population		
	Reference year	Issues
NL	2016	
BE	2016	
DE	2016	Two seperate not consistent data sets, one on total population and one on foreigners
DK	2016	
FR	2016	Citzenship imputed through calibration of the data broken down by sex, age and NUTS 3 of residence on the basis of 2014 results of census
IT	2016	Two seperate not consistent data sets, one on total population and one on foreigners
PL	2016	
SI	2016	

Although Luxemburg was not member of the consortium, we included data from Luxemburg in the analysis. The data was collected from the Internet from the websites of STATEC and the responsible ministries.

3.3. Results administrative data: cross-border workers

3.3.1. Cross-border workers per NUTS3 regions

Administrative data has been collected for NUTS 3 regions. This allows to include in principle all regions into the analysis. However, we would like to focus here on working in border regions and not to cover all persons crossing borders for work. In particular in large countries if workers that are resident in a bordering country that is positioned far from the region at hand they are not included in the analysis as incoming cross-border workers. This was applied to Germany and Poland. Concretely, for regions in the west of Germany, workers from Poland are not considered as cross-border workers. And for regions in the west of Poland, workers from Ukraine and Belarus were excluded. For all other (bordering) countries this was not necessary since that situation is uncommon and involve only a few workers. Note that for Italy country of residence was not available only country of residence 'other than Italy' was known. We took all persons with country of residence other and citizenship of neighbouring country of Italy. This will undoubtedly lead to an overestimation of the number of cross-border workers. Since not all Italians not living in Italy will be resident in a bordering country.

Map 3.1 shows the share of employees working in a NUTS 3 region that are resident in a border country for the countries DK, DE, NL, BE and FR. In case of several bordering countries the numbers are the total of all bordering countries. Only in a few regions cross-border work is significant. Luxemburg is with a share of more than 40% by far the region with the highest share of incoming cross-border workers. Other regions with relatively high shares of 5 percent or more incoming cross-border workers are some regions in BE bordering FR or LU, in NL bordering BE and DE, in the German federal states of Saarland (Saarbrücken and Saarlouis) and Baden Württemberg bordering FR (Rastatt) and in the German federal states of Bavaria bordering Czech Republic (Tirschenreuth and Cham).

Map 3.1. Share of employees resident in border countries for FR/BE/NL/DE/DK, 2015/2016
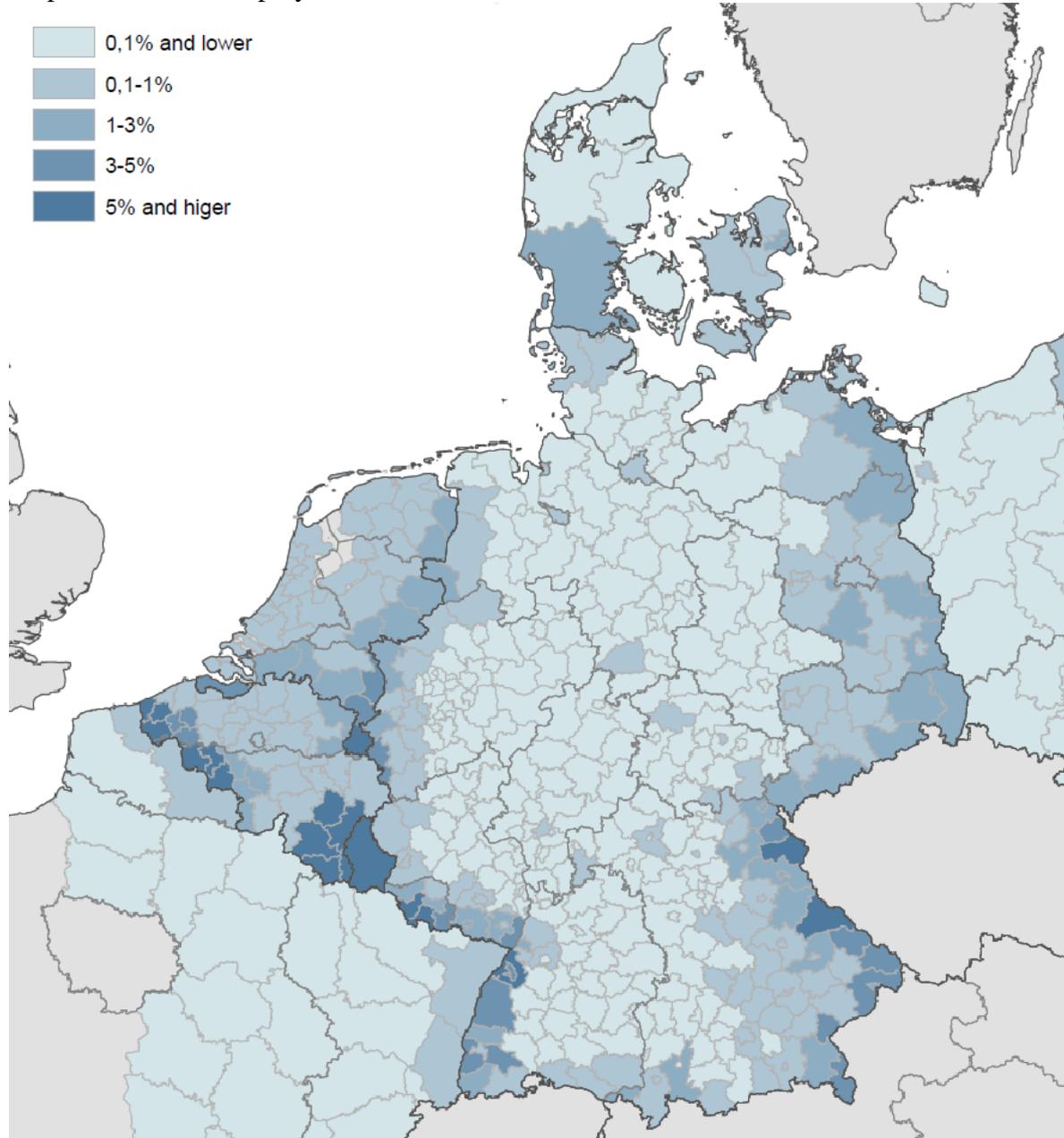

In other border regions cross-border work is much less significant. In the DK/SE border around Copenhagen cross-border work is relatively high. And at the Danish south side with the German border this is also the case. In the FR-IT border, cross-border work is very moderate with low shares. In the SI/IT border around Trieste there is some cross-border work activity. Finally, In Poland a moderate share of workers from Ukraine and Belarus work all across Poland. This does not show significant shares of incoming cross-border workers.

Map 3.2. Share of employees resident in border countries for FR/IT and IT/SI, 2015/2016
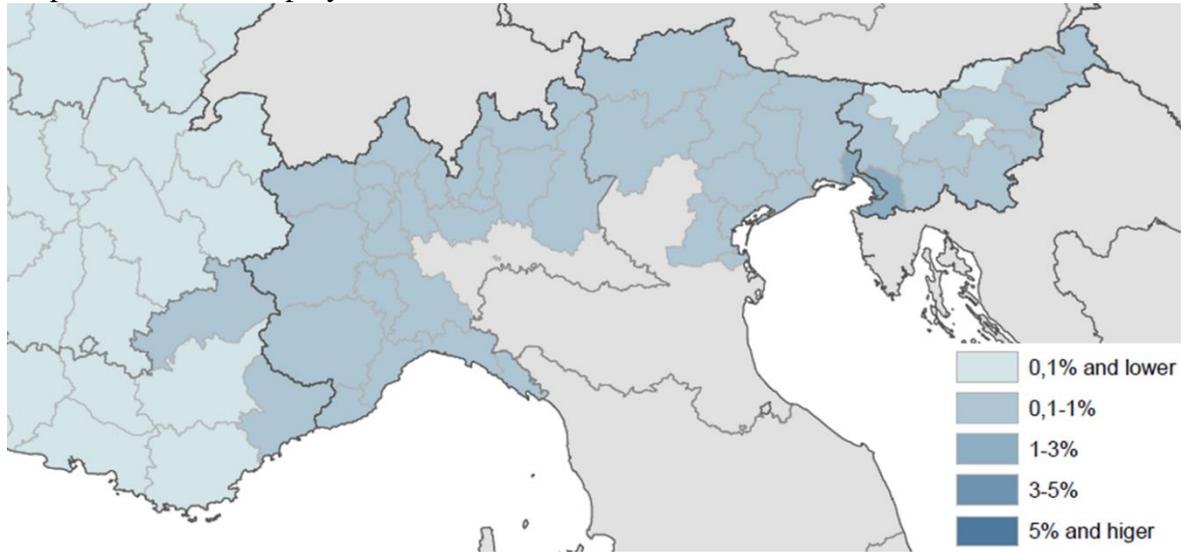
Map 3.3. Share of employees resident in border countries for Poland, 2015/2016
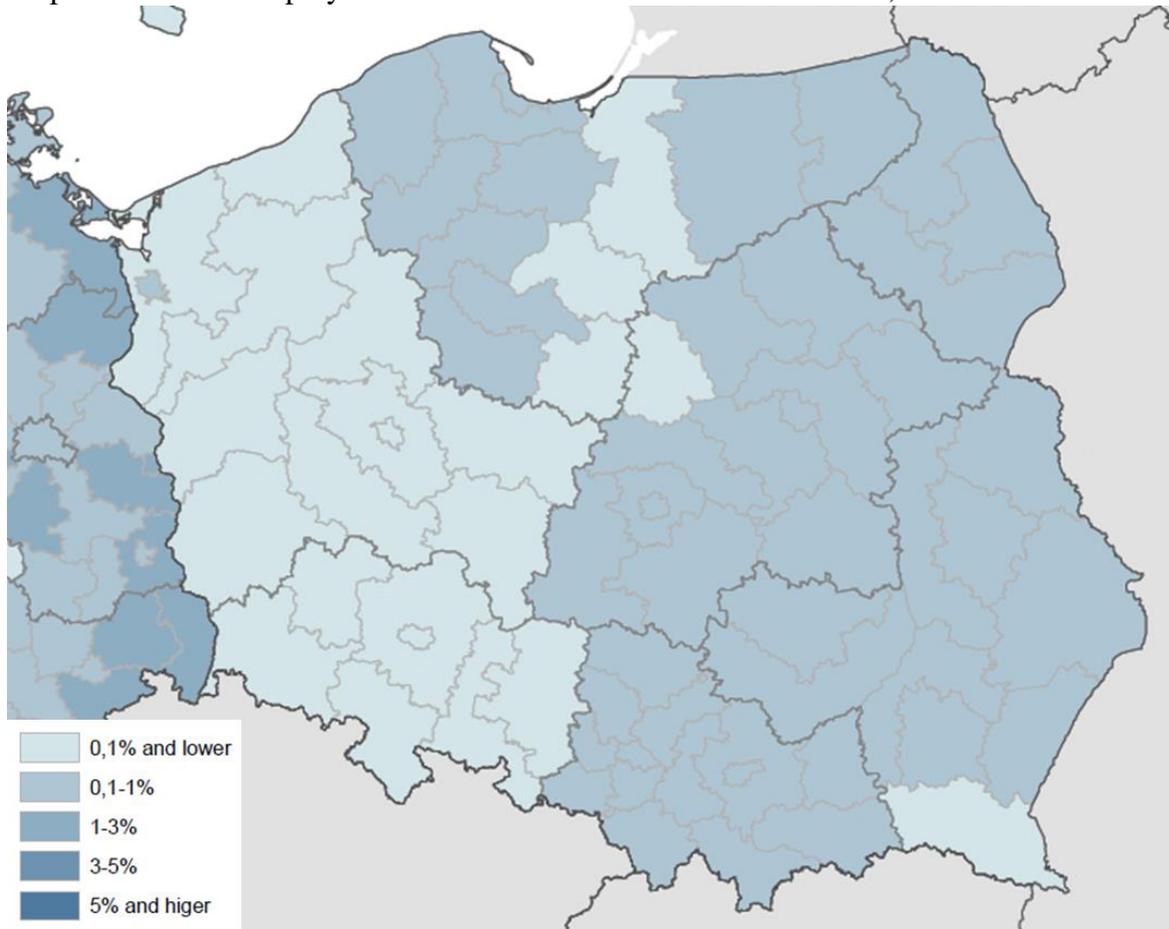
3.3.2. Cross-border workers by country of residence and citizenship

To show the potential is of the data collected we analysed the border regions more in detail looking at the country of residence and citizenship. For this we selected all regions with substantial shares of incoming cross-border workers with 1 percent or more. Since citizenship was not available for Belgium and France, these countries were excluded from the analysis.

The combination of country of residence and citizenship is quite complicated to understand. For this reason it is useful to explain the different categories of border workers. Note that we only consider incoming cross-border workers. So, they all work in a specific country, let us say country A. Cross-border workers have as country of residence the border country. The 'classic' cross-border workers are those with citizenship of the bordering country. But another relevant category are nationals living in the bordering country. They are country A citizens living in a border country but working in their home country. A third category, exotic and therefore generally small are persons with other citizenship than country A or the border country citizenship resident in the border country. Together these categories amount to the cross-border workers. In the context of cross-border work and cohesion there is an additional category of workers. They are workers with border country citizenship working and living in country A. The majority of them probably have migrated to the country of work. Scheme 3.1 presents all these categories. In the analysis we have identified all four categories as described above. Per border region we show the relative size of all these categories. It allows us to show how a standardised set of well-chosen indicators can reveal to what extent border regions differ from each other.

Scheme 3.1. Categories of incoming cross-border workers
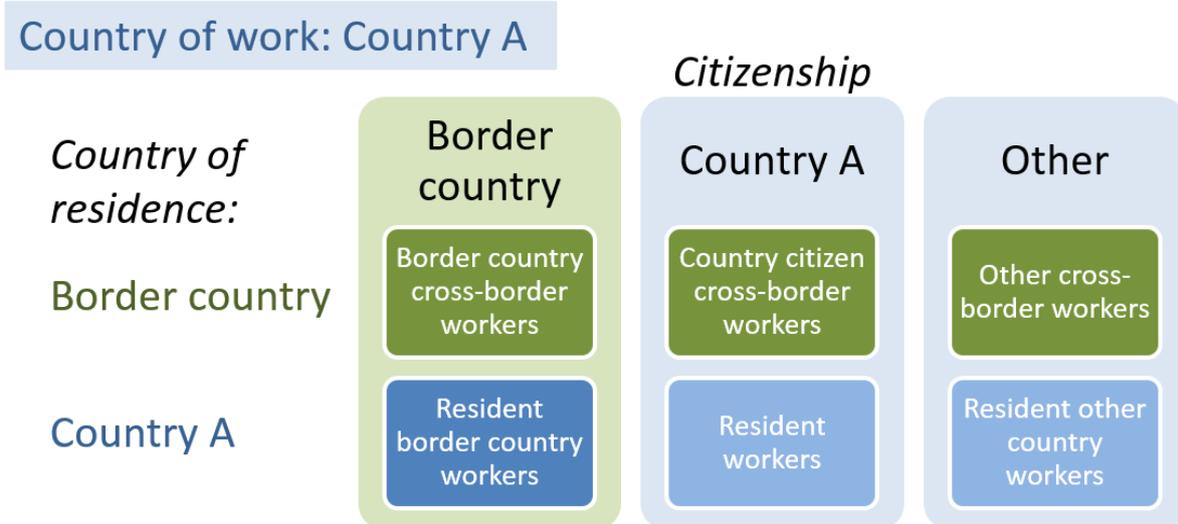
The first country we looked at is Denmark. Three Danish regions have share of 1 percent of more incoming cross-border workers around Copenhagen with the Swedish borders and Sydjyland in the south bordering with Germany. Copenhagen attracts Swedes coming from Sweden but also a significant share of cross-border workers are Danish living in Sweden. For the region Kobenhaven omegn this amount to half of the cross-border workers. At the German side most of the cross-border workers into Denmark are

Germans. Shares of persons with Swedish or German citizenship resident in Denmark are lower than the shares of cross-border workers with the same citizenship in their respective border regions.

Figure 3.1. Incoming cross-border workers by citizenship, Danish border regions, 2015/2016
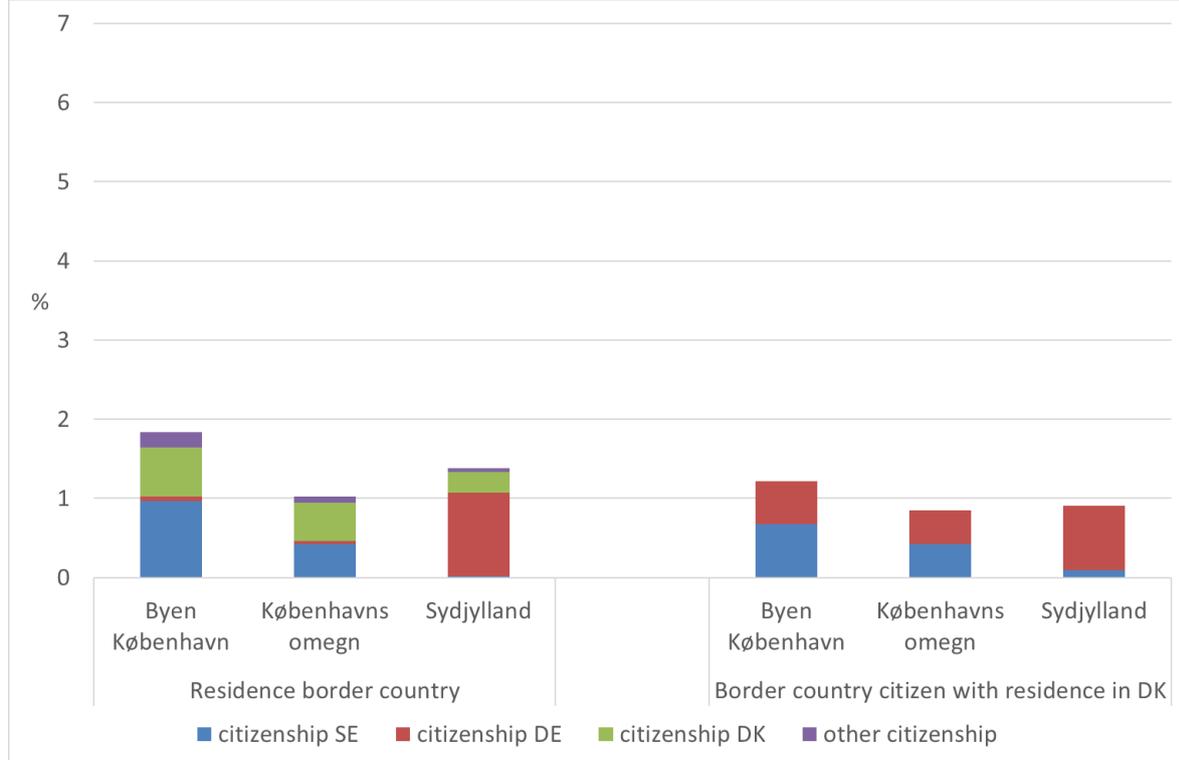
The border between Poland and German shows a quite straightforward pattern regarding cross-border work. Cross-border workers in Germany coming from Poland all have Polish citizenship for all regions. In the region Bautzen and Görlitz some of them are Czech, but this is logical because these regions border with the Czech Republic. Very little Polish workers are resident in the German border regions.

Figure 3.2. Incoming cross-border workers by citizenship, DE-PL/CZ border regions, 2016
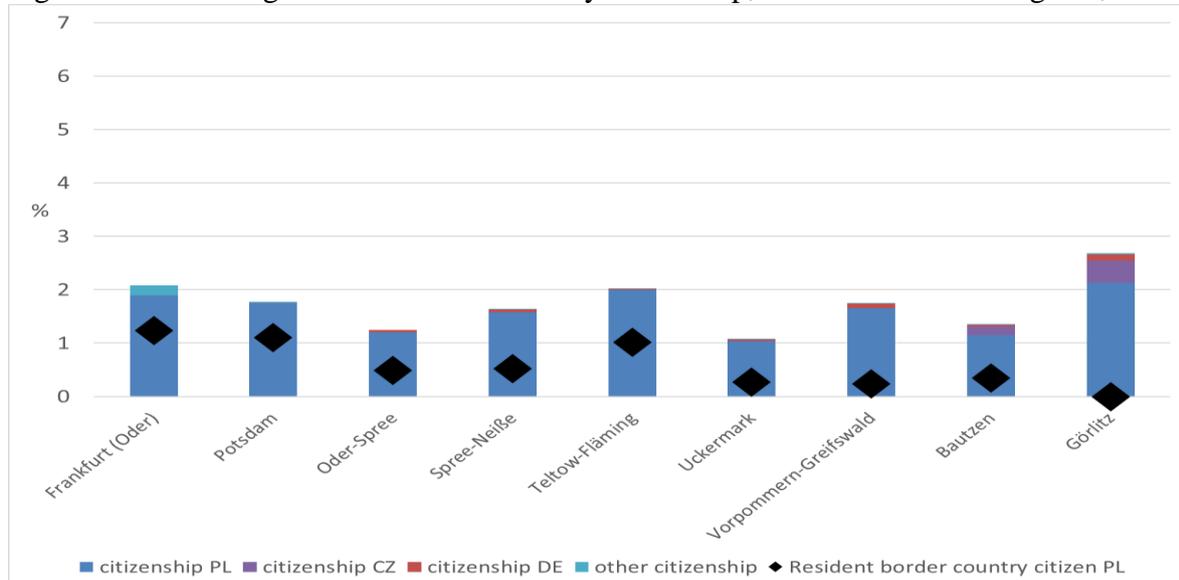
Looking at the German border with the Czech Republic, the situation is quite similar to situation at the DE-PL border. The shares of incoming cross-border workers are relative high for some regions. They are all Czech citizens. Except for Passau that borders with Austria as well, were a significant number are Austrians. Czech citizens do not live in the German border regions. This category is non-existent at that side of the border.

Figure 3.3. Incoming cross-border workers by citizenship, DE-CZ/AT border regions, 2016
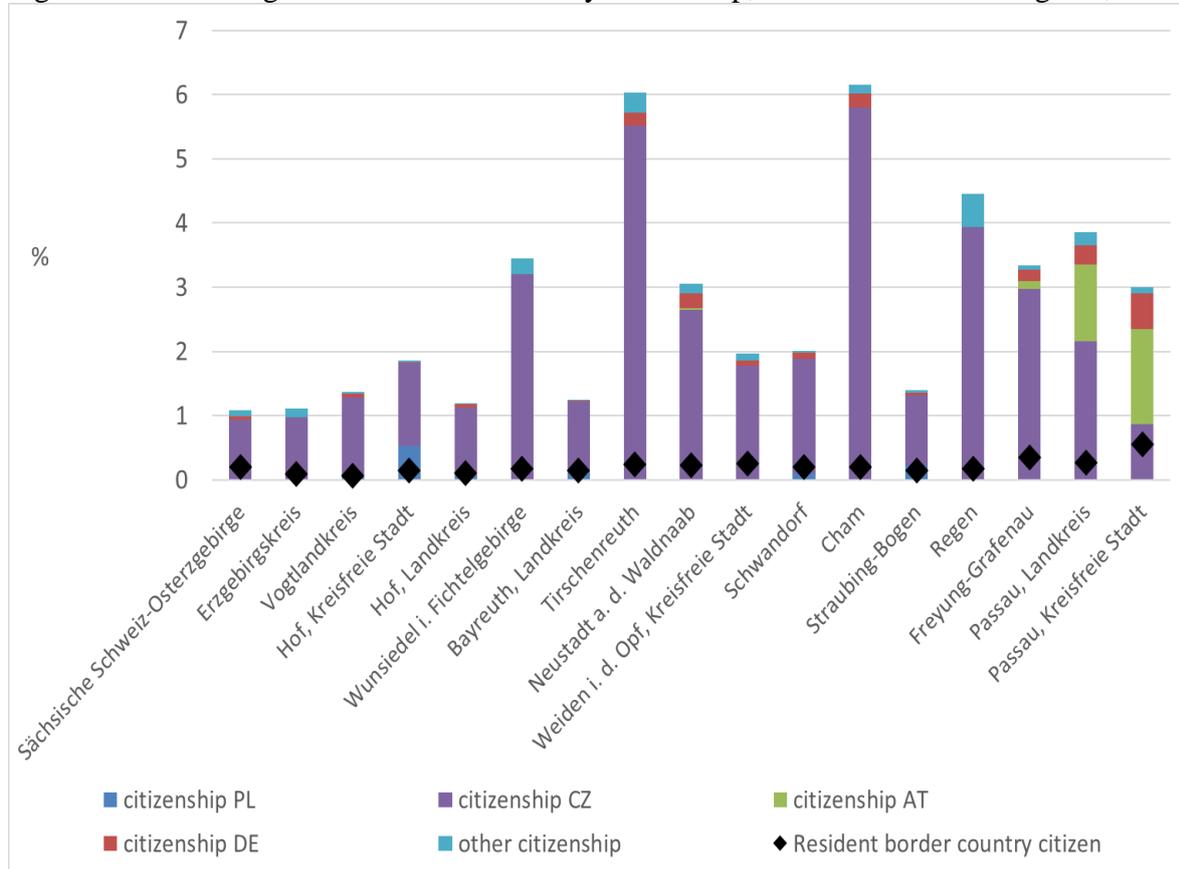
The German-Austrian border shows a bit less straightforward situation regarding cross-border work. The vast majority of the persons working in the German border regions while living in Austria are Austrians. But there are also some German cross-border workers. One of every five has German citizenship. Austrians that live and work in German border regions are virtually non-existing.

Figure 3.4. Incoming cross-border workers by citizenship, DE-AT/CH border regions, 2016
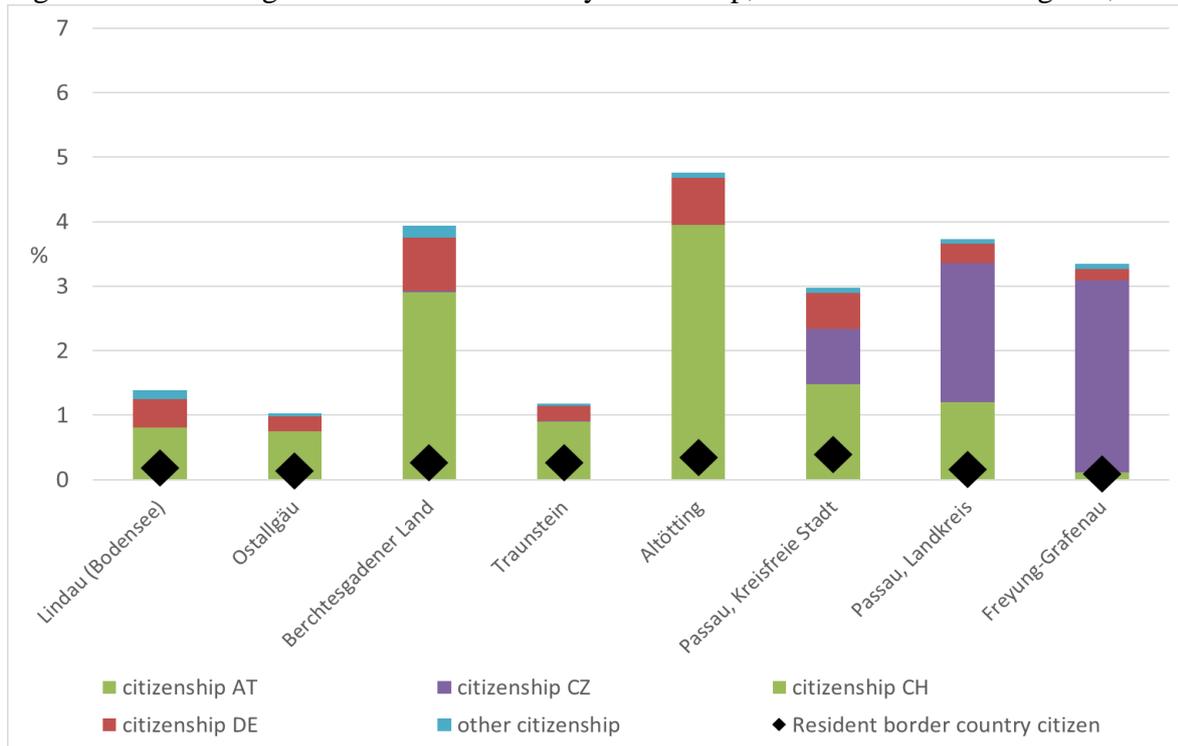
The French-German border shows a number of regions with high shares of 5 percent or more incoming cross-border workers: Saarbrücken, Saarlouis in Saarland and Baden-Baden and Rastatt in Baden-Württemberg. In all regions the majority consists of French citizens. But also a substantial share are Germans living in France. This amounts to about a bit less than a quarter of the cross-border workers in these regions. French workers do not live in Germany when they work in those regions.

Figure 3.5. Incoming cross-border workers by citizenship, DE-FR border regions, 2016
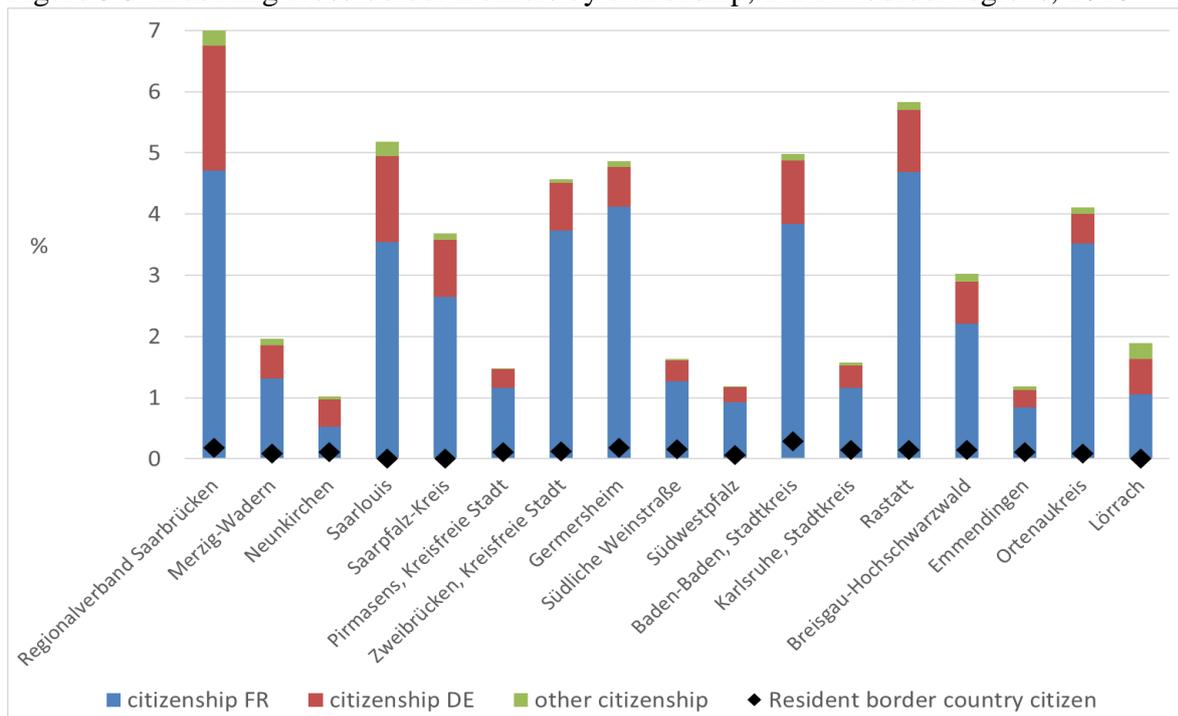
Now we turn to the more complicated regions. First we look at Luxemburg, champion of cross-border work. Since Luxemburg is surrounded by three larger countries, France, Belgium and Germany we need to identify all those countries for country of residence and citizenship. Half of the incoming cross-border worker are from France: 22 percent. The other half is coming from Belgium or Germany split evenly: 11 percent both. Virtually all cross-border workers are of the 'classic' nature with citizenship of the resident country. Although one observe some numbers of Luxembourgish workers living in the neighbouring countries, in total these numbers are small. This is not surprising since Luxemburg is small compared to its neighbours. In Luxemburg resident workers from neighbouring countries are limited. But still French living and working in Luxemburg amounts to 4 percent of total employees.

Figure 3.6. Incoming cross-border workers by citizenship, LU border regions, 2016
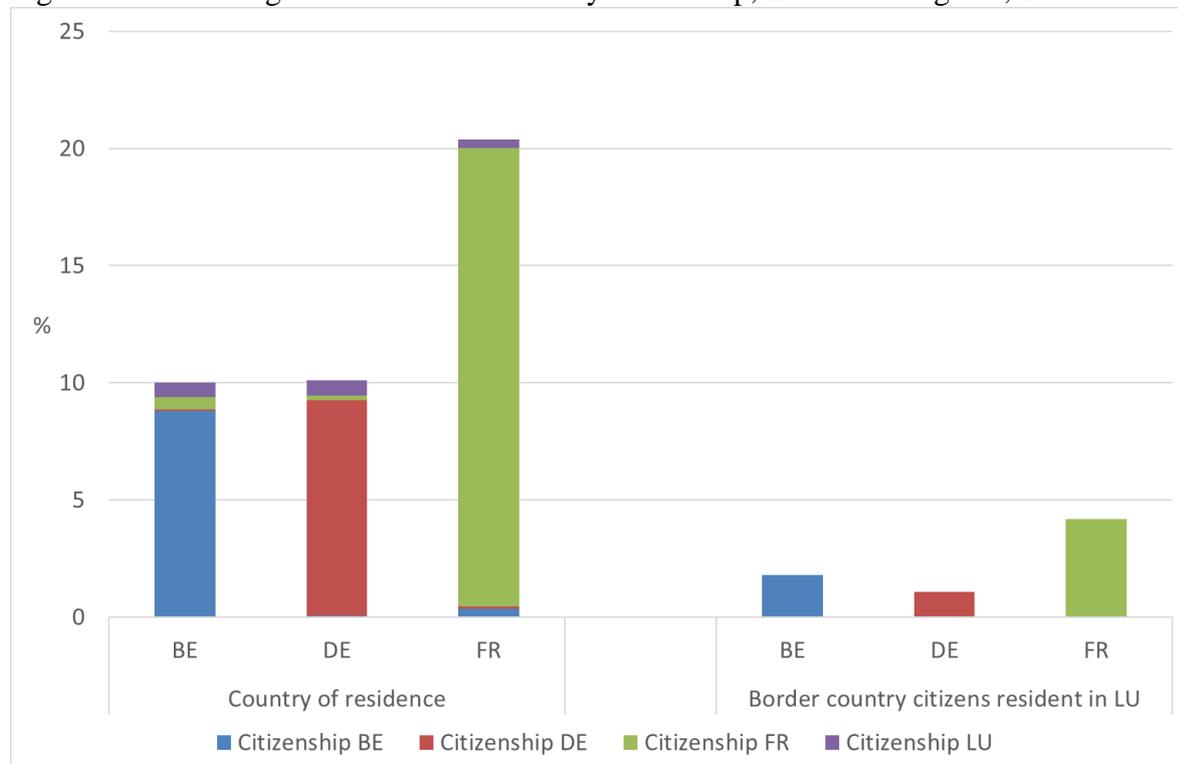
Although the Netherlands is not a large country quite a number of regions show 1 percent or more shares of cross-border workers. The Netherlands shares borders with Germany and Belgium. As a consequence the situation varies quite a lot between the border regions. The regions in the north part of the Netherlands bordering Germany show moderate shares of incoming cross-border workers. In all those Dutch regions the majority of the cross-border workers are Dutch. Germans are hardly crossing the borders in those regions. The two German regions in that area with shares of one percent or more are Grafschaft Bentheim and Kleve. Those regions also show Dutch workers crossing the borders being 'classical' cross-border workers. The region Noord-Limburg also bordering Germany shows a different picture. There the majority consists of workers with German citizenship. In Midden-Limburg and Zuid-Limburg, bordering

with both Germany and Belgium about half of the cross-border workers are Dutch living in either Germany or Belgium. The German region of Aachen is quite unique in the sense that more than half of the incoming cross-border workers are Germans living in Belgium or the Netherlands. To sum up the description of the Dutch-German border one can observe relative high numbers of Dutch citizens working and living in the German border regions while hardly any German citizens living and working in the Netherlands.

Figure 3.7. Incoming cross-border workers by citizenship, DE-NL border regions, 2015/2016
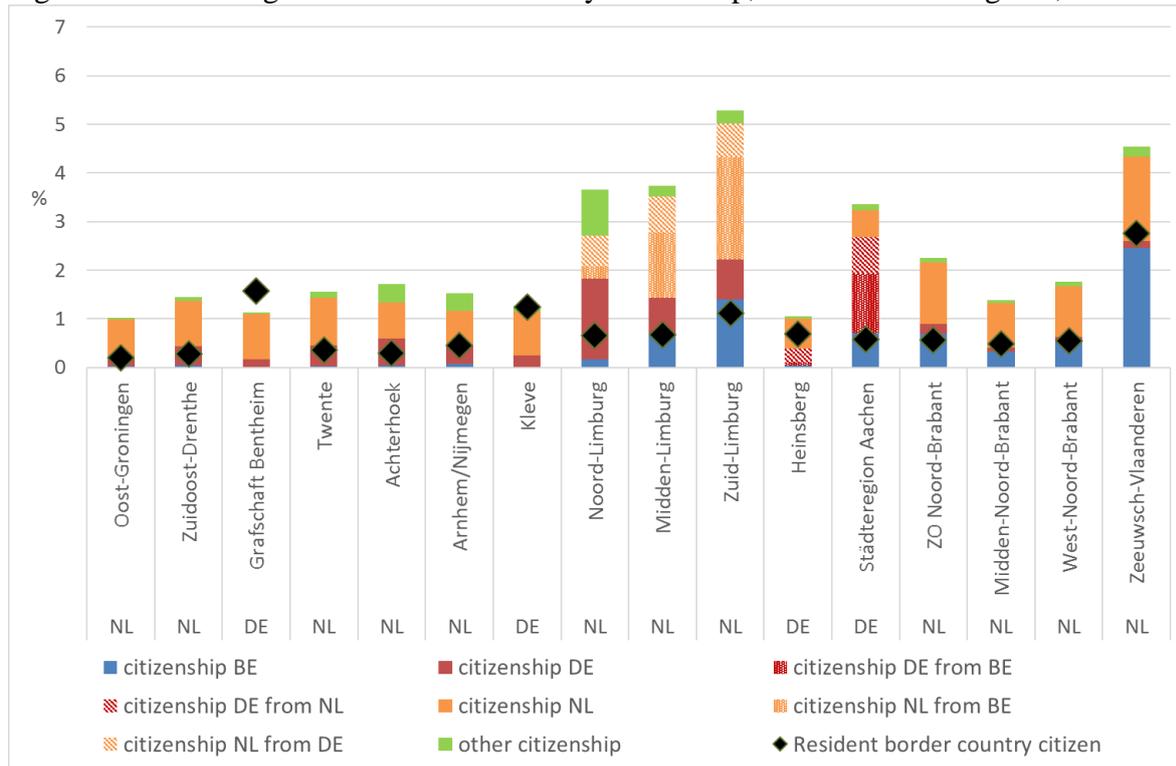
Looking at the Dutch-Belgian borders one sees a similar diverse picture. The three Dutch border regions of Noord-Brabant border with Belgium. In those regions the majority of the cross-border workers consists of Dutch persons living in Belgium. Zeeuws-Vlaanderen is more or less an enclave in Belgium sharing only a land border. As a consequence, the share of cross-border work with Belgium is considerable. About half of them are Belgian cross-border workers and half of them are Dutch persons living in Belgium. Also the share of Belgian people living and working in the Netherlands is considerable.

Finally, we look at the Italian-Slovenian border. Only two Italian and one Slovenian region shows shares of 1 percent or more cross-border workers. They are all regions in the Trieste area. At the Italian side relative large part consist of Italians living abroad but working in Gorizia or Trieste. The classical cross-border workers with Slovenian citizenship is relatively small. In the Slovenian region of Obalno-Kraska the majority is also Italian. Resident border country workers are virtually non-existent in the Italian regions. In the Slovenian region live and work quite a number of Hungarians and some Italians.

Figure 3.8. Incoming cross-border workers by citizenship, SI-IT border regions, 2015/2016
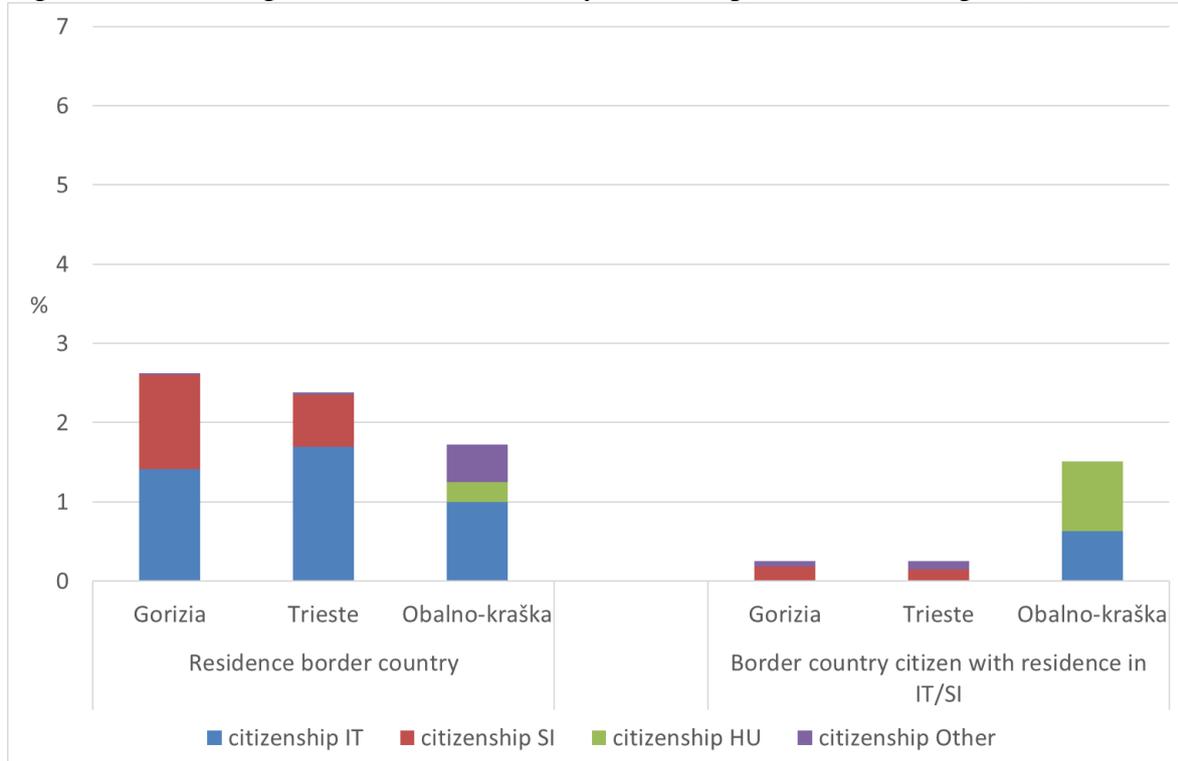
3.3.3. Cross-border workers by sex

Another characteristic that was included in the data file as background information is sex. It is interesting to see if are differences between men and women regarding cross-border work. Unfortunately, no data for Luxemburg was available. For the other countries we used the data for all the NUTS-3 regions available. For France and Italy this were all NUTS 3 regions that are part of a NUTS 1 region close to the border. For the other countries all regions were included.

The results are very clear. For all countries that was included in the study, the share of incoming cross-border workers is higher for men than for women. It suggests that men are more likely to cross the border for work than women. In Belgium, Germany, Italy the shares of incoming cross-border workers are for men about twice as high as for women. In Denmark, France and the Netherlands the differences between men and women are somewhat smaller. For Poland and Slovenia, these differences are considerably higher.

Figure 3.9. Share of incoming cross-border workers by sex, 2015/2016
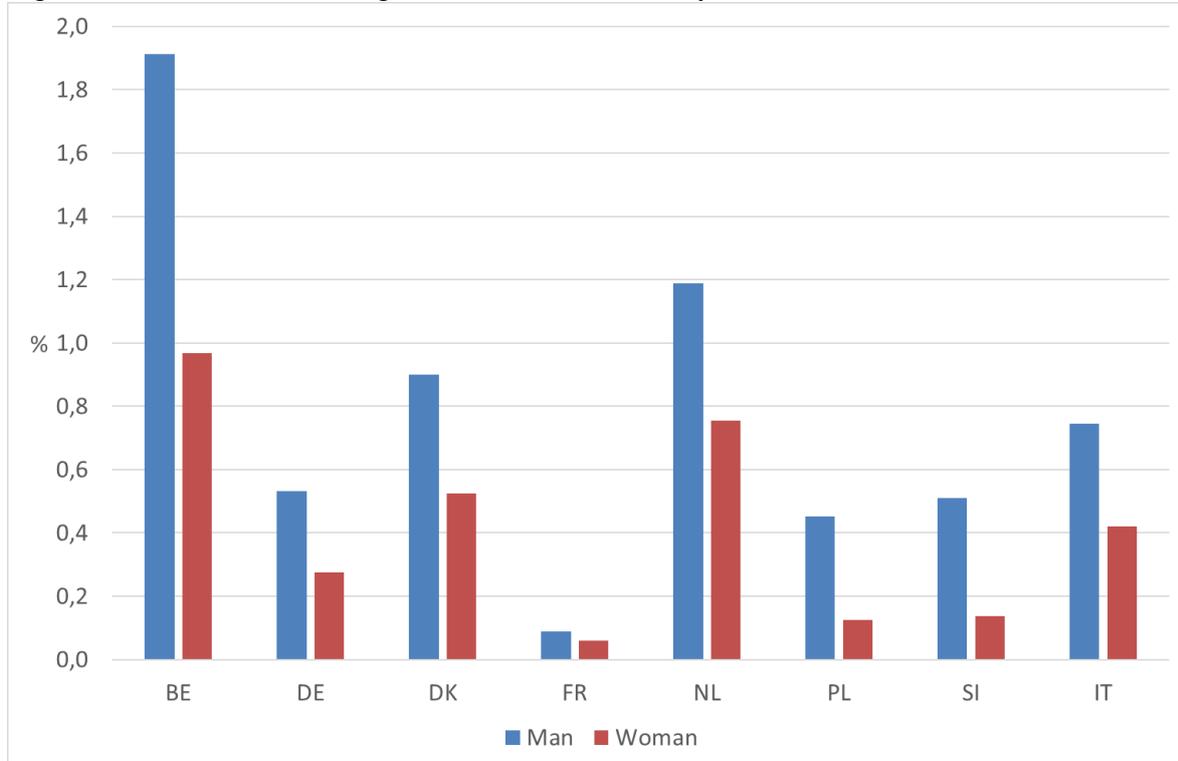
3.3.4 More potential of administrative data: the Danish case

The Nordic countries have a long standing tradition in using administrative data for making statistics. They have developed very sophisticated applications of the use of administrative data. For this reason we were very happy that with Denmark we had a representative of this community. We asked our Danish colleagues to show what is possible to do with administrative data in a cross-border context. Concretely we asked for grid data on cross-border workers. To show what is possible they produced some graphs for both the border with Sweden as the border with Germany. In Denmark, they are able to use a 100mx100m grid to provide data. The results for the Copenhagen areas that is connected with Sweden with Öresund bridge are presented in this report.

The map clearly show that incoming cross-border workers are focussed in Copenhagen which makes sense. Interesting to see is that around Copenhagen the Swedish workers are not spread evenly in the areas. But one see two strips west of Copenhagen. They are along the two main motorways and railways connecting Copenhagen with the country side. It seems that for cross-border commuters the time to get to work is an important factor. Good connections can make a difference apparently. This information is very relevant in a cross-border context.

Map 3.4. Share of incoming cross-border workers from Sweden per 1kmx1km grid, 2015
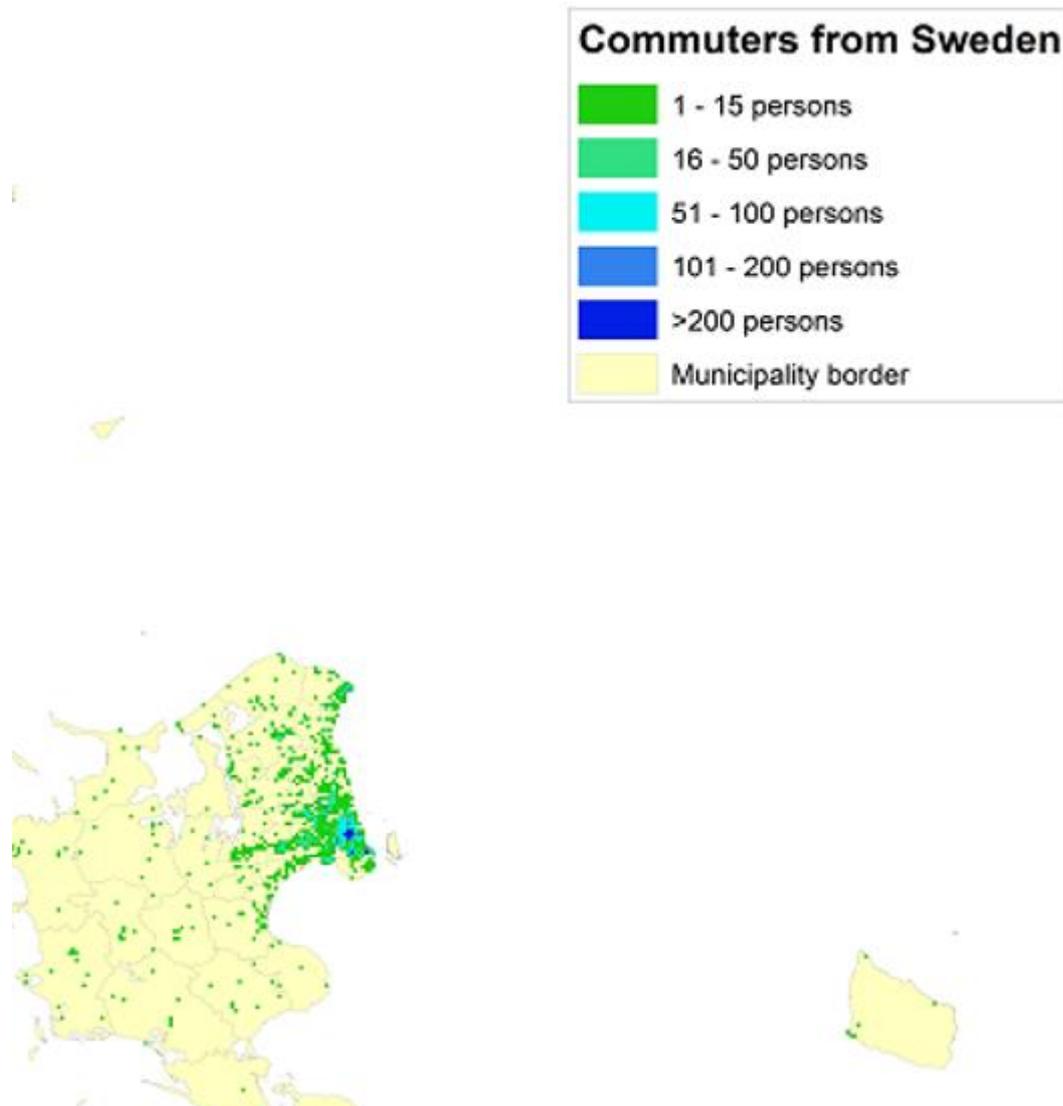
Since data is available for a 100mx100m grid it is possible to zoom in. Below is a detailed map of the city of Copenhagen pointing out cross-border workers from Sweden. One can clearly see the main hotspots of work and the shopping streets in the city centre. Were many people work one will find also more cross-border workers. This is not very surprising. It seem also less relevant from a cross-border perspective. On can conclude that this level of detail within a city is maybe less needed in a cross-border context.

Map 3.5. Share of incoming cross-border workers from Sweden per grid in detail, 2015
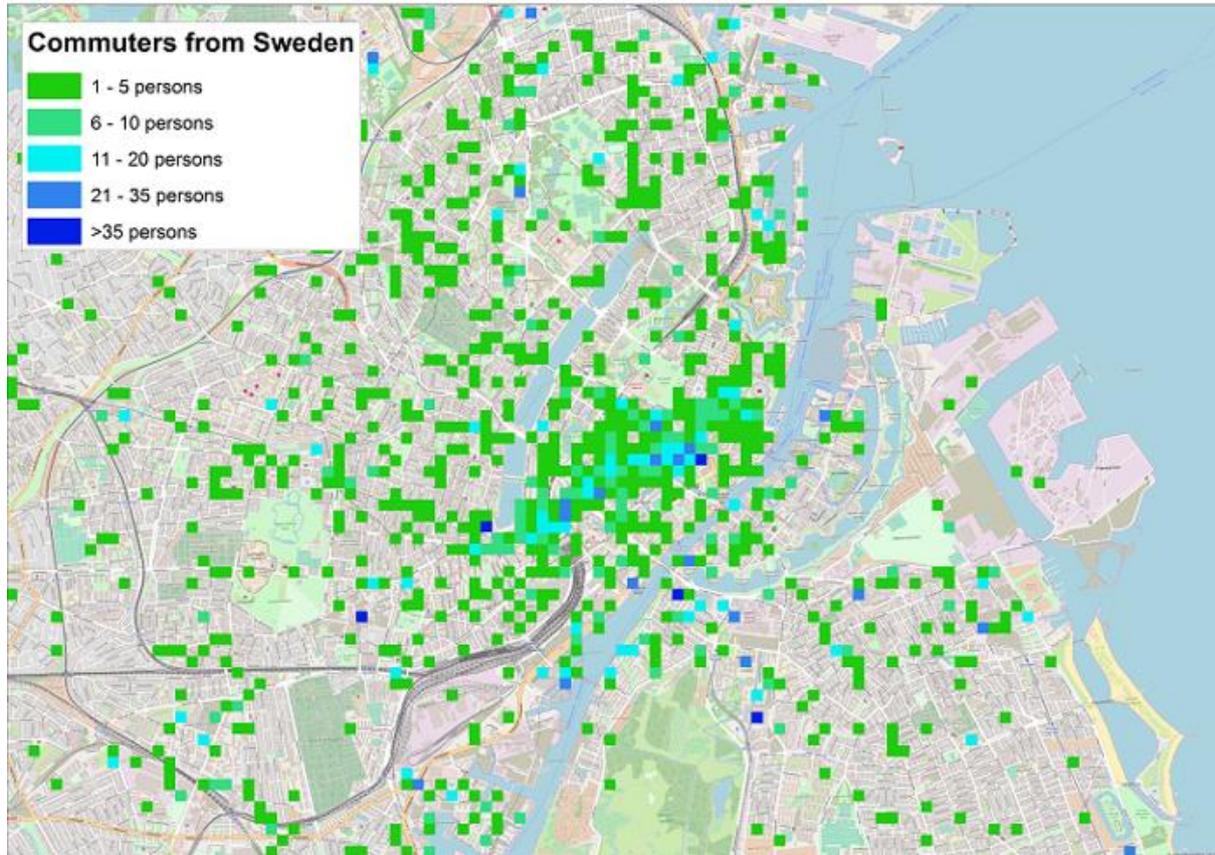
3.4. Results administrative data: population in border regions

3.4.1. Population in border regions: elderly

Generally, high shares of elderly is considered as a sign of more vulnerability. Therefore, it could be interesting to see if border regions show high levels of elderly. The share of 65+ shows no clear and structural picture in border regions. Some regions show high shares of elderly while other show relatively low shares. It also differs strongly per country. Denmark, Italy, Belgium, the south of France and the east of Germany shows generally high shares while Poland shows low shares. In Germany, the eastern and northern border regions show high share. This also the case in some Dutch border regions. In contrast, French border regions with Belgium, Luxemburg and Germany do not show high shares.

Map 3.6. Share of population 65 years or older, 2016
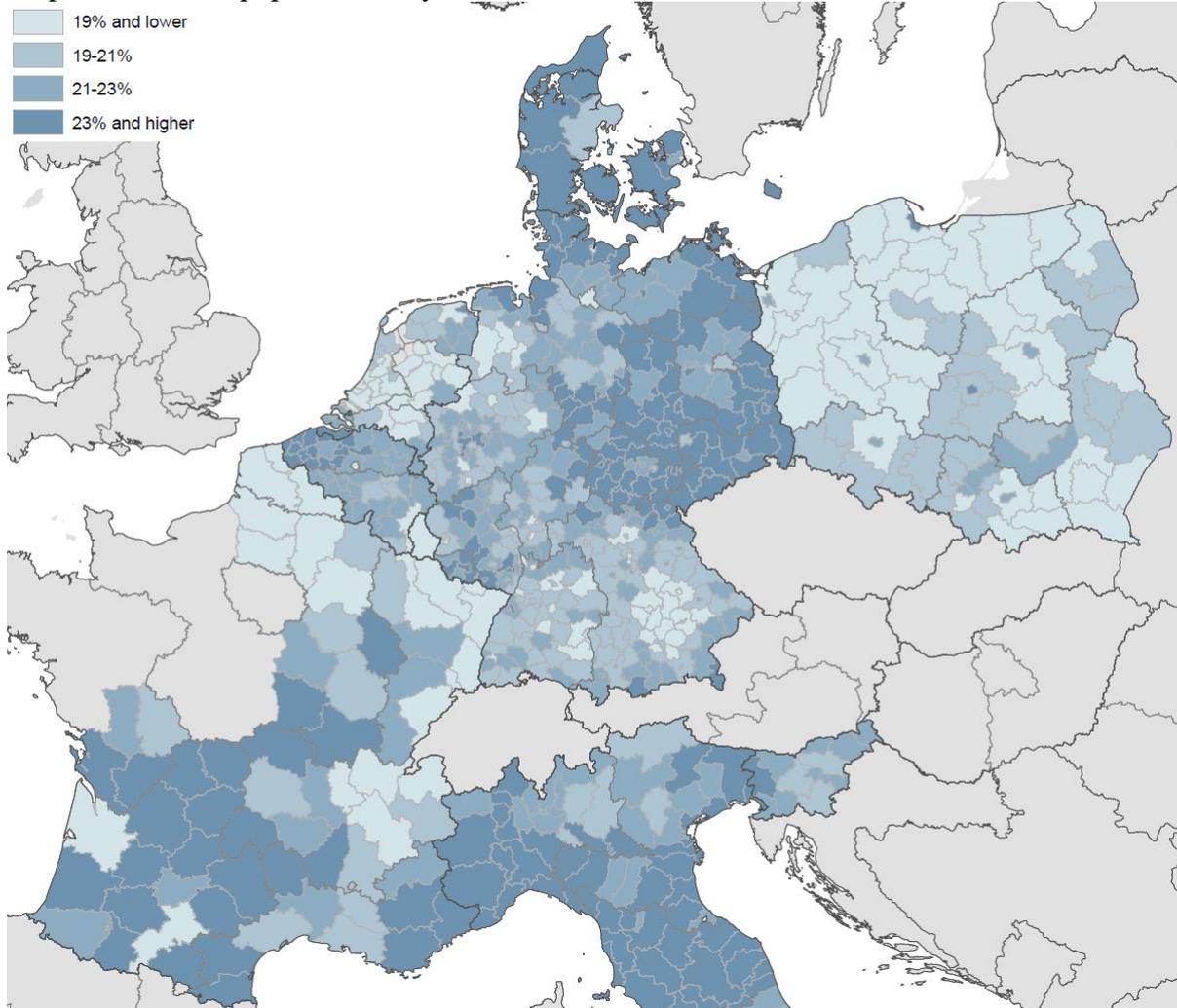
More analysis is needed to answer the question what extent this information is relevant and useful to explain differences. Within the time frame of this project we were not able to go into depth with this data. But our initial opinion is that it is from a cross-perfective only of limited interest. It is not an issue that is relevant to border regions specifically but is also relevant to non-border regions.

3.4.2. Population in border regions: share of border region citizens

The population data on foreigners is not without issues. French data is from the 2014 results of the Census. Both Germany and Italy have two distinct not consistent administrative data sets on the population and on foreigners. As a consequence, for these countries we cannot calculate the share of persons with border country citizenship of the total population. Looking at the other countries, we can see that in Denmark

Copenhagen has a large share of border country citizens. The same is the case in several regions at the NL-BE, FR-BE, BE-LU, BE-DE and FR-DE borders. The regions that are prominent in Luxemburg, Belgium, the Netherlands fits quite well to the results of the analysis of the data on cross-border workers. One would namely expect that high shares of resident border country workers in the total number of employees lead to high shares in the population and incoming cross-border workers with the country citizenship corresponds with high shares of country border citizens in the regions just across the border. In Poland no clear picture regarding border regions is visible. For Slovenia the southern border regions show high share of border country citizens.

Map 3.7. Share of foreign population of the direct bordering countries in the total population, 2016

|	
	0,1% and lower
	0,1-0,3%
	0,3-0,6%
	0,6-1,5%
	1,5% and higher
	missing data
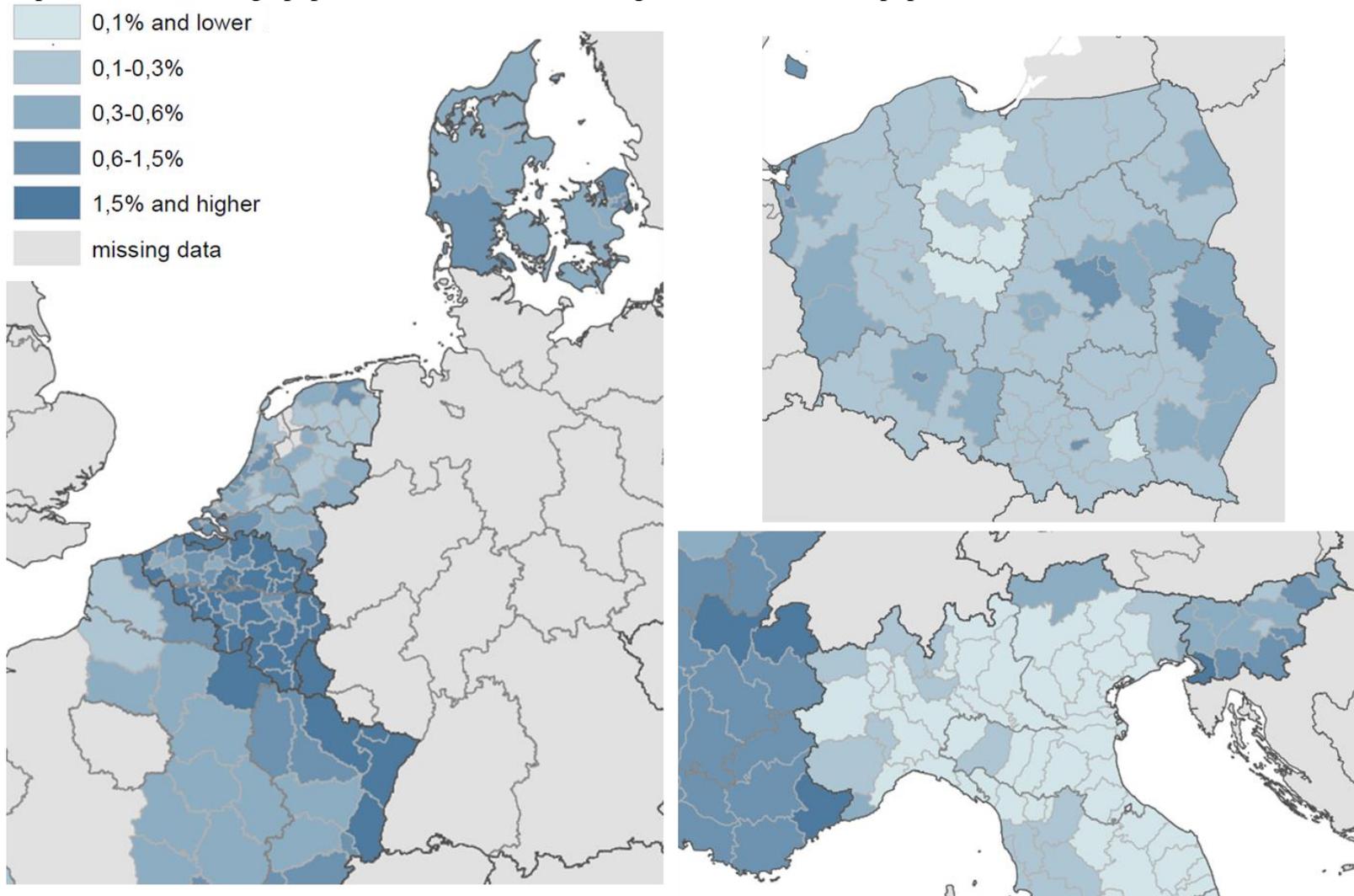
An alternative indicator is the share of border country citizens within the foreigners. This can be produced for Germany being a crucial country for analysing cross-border issues. We would like to know if this indicator is useful and what the added value is compared to the more straightforward indicator of border country citizens as share of the total population. Theoretically, this second indicator is less valuable to measure what we want to capture, namely how many persons settle in their neighbouring countries. It is also less useful to compare regions because it depends on the presence of other foreigners. If this differs between regions then this second indicator will be distorted.

To see what the added value and the usefulness is of this indicator compared to the share of the total population in practice we looked at a specific region. We took the case of the Benelux with their external borders since this is an important region from a cross-border perspective. The share of border country citizens of the total population shows that almost all border regions in Belgium show high shares, as well as Luxemburg, Zeeuws-Vlaanderen and Zuid-Limburg in the Netherlands and the north-east borders of France. These results correspond with the results of the data on cross-border workers. Remember that we noticed that the share of incoming Dutch cross-border workers in the south of the Netherlands is high which is a sign that they live across the border in Belgium. In addition, we found a significant number of Belgian people living in Zeeuws-Vlaanderen and Zuid-Limburg. Similarly this is the case to a somewhat lesser extent at the FR-DE border with limited numbers of Germans living in France and working in Germany. In Luxemburg we say a substantial share of French citizens working and living in Luxemburg, which fits as well.

Let us now look at the indicator of the share of border country citizens within all foreigners. First, the positive side: data for Germany. We see that at the DE-NL borders two regions stand out: Grafschaft Bentheim and Kleve. This corresponds well with the data on cross-border workers. In those regions relatively high shares of Dutch people are living there and substantial numbers of Dutch work in the regions across the border in the Netherlands while living in the Netherlands. This is quite understandable if you look at the geographical location of the regions. Also the regions that now stand out in Belgium are the regions you would expect. The regions Turnhout, Maaseik , Tongeren host many Dutch people, the regions
Mouscron, Tournai, Virton is popular with French people and the Deutschsprachige Gemeinschaft with Germans. So this is plausible and useful.

On the negative side, one has to conclude that North-Limburg Zuid-Limburg in the Netherlands and Luxemburg now seem somewhat less important from a cross-border perspective than one would expect. This can be explained because of the presence of substantial numbers of other foreigners apart from those of the border countries. This is probably also the reason why these regions are now less prominent. This is maybe also the reason why the French regions are less prominent.

Map 3.8. Share of foreign population in total population, 2016

	0,1% and lower
	0,1-0,3%
	0,3-0,6%
	0,6-1,5%
	1,5% and higher
	missing data
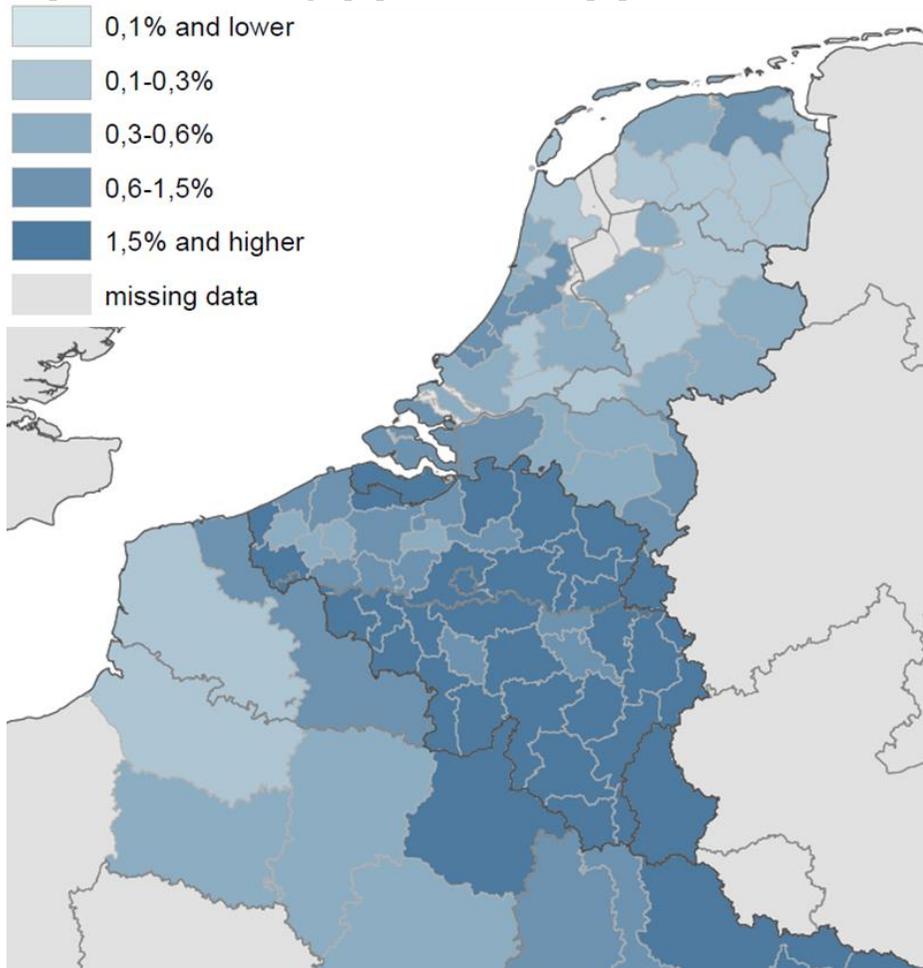
Map 3.9. Share of foreign population in total foreign population, 2016

	7% and lower
	7-17%
	17-27%
	27-50%
	50% and higher
	missing data
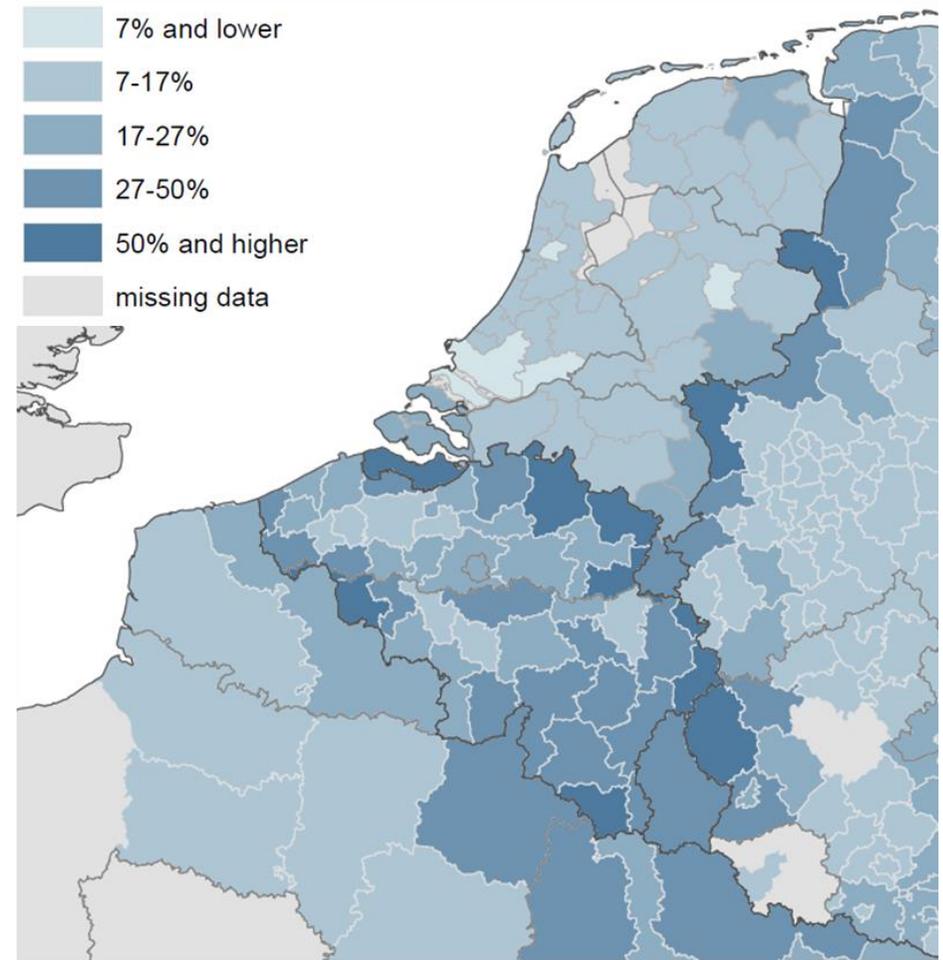
Concluding based on this analysis, one could say that the share of border country citizens gives plausible and useful information. But it should be handled with caution when comparing regions because of differences in shares of foreigners per region. It also fair to conclude that the insights that are gained from population data does not add a lot of information of what already can be determined based on the cross-border workers data. The added value of population data is limited.

3.5. Conclusions regarding administrative data

First we can conclude that collecting administrative data is not straightforward. Firstly, all countries have different administrative systems with different sets of data. As a consequence, some information could be available in one country but not necessarily in another country. An example of this is citizenship that is not available in France. Secondly, the comparability of the data is not guaranteed. It can be the case that some specific sub-populations are not covered (like civil servants or the agricultural sector). It can also be that some information is of less quality in some countries than in others. A good example of such a characteristic of the place of work in case of companies with several location or temporary workers. Thirdly, specific to cross-border data is that the population of persons not resident in the country is often excluded from the target population national statistics. To gather information for these cases, even in standard administrative file, is not business as usual. They are exotic and rare cases not very relevant in national context. It requires extra work to access the data and check the quality of the data. Finally, not all statistical institutes have integrated systems in place of combining or linking data. This is for instance the case in Germany and Belgium. Italy is still developing such a system.

We conclude that collecting data on incoming cross-border workers from social security files is feasible. In all countries this was possible with a limited effort.2 The results seem plausible. We have not done an in depth analysis of the quality and the comparability. There will be some issues if you look into the details. But overall we think that the results are usable. Moreover, the data is quite timely. Countries are able to provide data within 1-2 years after the reference year.

We were able to provide information at high spatial resolution (or at the level of small regions or territories) on incoming cross-border workers based on administrative data for all countries in the project. This is an extremely valuable result. That kind of data is currently not available at EU-level. This project gives strong evidence that it is possible to provide such data for most of the European countries. We have developed a relatively simple template for data at NUTS 3 level that can be used for such a data collection process.

We would like give some suggestions to extend the information that is collected from administrative files on cross-border work. First, the level of regional detail. In a number of countries NUTS 3 regions can be quite large. This is of course not very helpful to provide information on border regions. Several countries apply nationally smaller regions than NUTS 3 regions to provide labour market information. In France and in Italy they use Labour market areas (LMA's), see Annex II. We advise to use in those cases these

---

2 Belgian data was not collected within the time frame of the project but this data can be produced. Even citizenship would be available.

kind of regions rather than the NUTS 3 regions. In principle administrative data can be collected on lower regional level than NUTS 3 or LMA's. Data on LAU 2 level is in principle possible. The Danish example show that we can even go further and provide data on for grids. We advise to explore these approaches further.

The second suggestion is about the place of residence of incoming cross-border workers. This is a very important piece of information to have. We know where people are working in the host country but not where exactly they are living in the sending country. This kind of information was not readily available for most of the countries involved. Therefore, it was excluded from the data collection in this project. However, we know that in a number of countries can produce this information with some effort. This is the case for Luxemburg, the Nordic countries, the Netherlands and Germany. We suggest to start an initiative to acquire this information for the countries where this is possible.

Finally, we collected population data on border regions. This was not without issues because for instance in Germany no integrated data on foreigners and the resident population is available. As a consequence, the indicators are not straightforward to interpret. Do we really need such additional data? We think not. Firstly, population data is to reasonable extent already available at EU level provided by Eurostat. Let us use that. Secondly, data on cross-border workers including citizenship already gives similar information on which people live and work where. The added value of extra population data is limited. A cost-benefit analysis would in our opinion lead to the conclusion that it is nice to have but not essential.

4. MOBILE PHONE DATA

4.1. Introduction

Mobile Phone data seems to have high potential for production of cross-border statistics. Based on mobile phone records, so-called Call Detail Records (CDR), it is possible to estimate location of the devices on a low regional level in a certain moment. Moreover, CDR-data is a global standard used by all the mobile phone providers. What makes possible to produce statistics based on Mobile Phone data using the same methodology across the globe.

This chapter presents the work that was carried out in the project regarding mobile phone data. In this part of the project had a limited scope. Contrary to the other working packages for this working package only the Statistics Belgium and Statistics Netherlands participated. The other institutes had no access to mobile phone data at the time of applying for the project. Secondly, developing methodology had to start from scratch. At the time of starting the project it was not clear which data was available and how raw the data was. Since the timeline of the project was limited and the resources very limited only an initial part of developing statistics was possible. The object was to make progress in developing methodology and get a glimpse of what is possible in the future.

Another reason to keep the scope of the work on mobile phone data limited in this project was that the method should be well coordinated with the ESSnet on Big Data in which also mobile phone data is one of fields that is worked on. In the ESSnet-project the focus will be on internal country movements while this project concentrates on the special case of cross-border movements. To further ensure this coordination, the members of the mobile phone data WP of the ESSnet on Big Data of BE and NL were also involved in the Border region data collection.

4.2. Developing methods

Before the start of the project the initial idea was that Statistics Netherlands and Statistics Belgium would both get access to aggregated Mobile Phone data in given country. Statistics Belgium already had access to Belgian mobile phone data of a major provider at the time when the consortium applied for the project. Statistics Netherlands was in the process of acquiring data from one major provider in their country. Unfortunately, at the kick off meeting of the project the data availability situation in Belgium had changed completely. The Belgian data provider had recently terminated the cooperation with Statistics Belgium and would not allow to use the old data. So Belgian data was not available. Therefore we could use Dutch data only. The Belgian colleagues and expertise was used in the project as input for and assistance with developing methodology.

At the Dutch side the process of establishing a cooperation with mobile phone provider and subsequently getting access of the data took longer than anticipated. This process took so much time that we were practically half way within the time line of the project. It is important to mention that Statistics Netherlands did not get access to the original raw micro data but received only aggregated data. The methods for aggregation were

developed in collaboration between statistical institute and the mobile phone provider. In this way the whole data process is transparent.

With CDR data it is possible to measure time and location of events of mobile phone devices. The location is derived from the antennas the device made connection with. In addition, the country code of the device is recorded. This is essential for cross-border information. Most of the timing of the project was dedicated to acquire a test set of data that included this information. In particular to develop an satisfactory measurement of the location was time and resource consuming. The method that we applied uses the location of the antenna(s) and the strength of the signal in a Bayesian model. This method is described in section 3.2.3 in the paper delivered in working package 5.3. in the ESSnet project on Big data (to be found here). This paper including the other papers within the same working package on mobile phone data provides much more relevant information on the state of art of using mobile phone data for statistics.

By the end of the project we were able produce some promising results. The results of a data set of a period of several weeks on Belgium and German devices showed time and location patterns that were considered very plausible. We looked at timing of the number of German and Belgian devices per hour, days within the week and weeks within the month for a few selected municipalities. This revealed highly recognisable patterns. Border towns clearly showed up prominently as 'hot spots' for presences of border country devices. Numbers were generally low at night time, somewhat higher at weekend and showed a certain pattern on weekdays.. In addition, it showed we saw that the German and Belgian Easter holidays showed up in the period we looked at. We know that the Netherlands is a popular destination to visit during these holidays. The slightly different timing of these holidays per country was clearly reflected in the data. Figure 5.1 and figures 5.2 presents examples from patterns that were visible in a period in the spring of 2018.

Figure 5.1. Pattern of number of devices in time of a Dutch town at the Belgian border
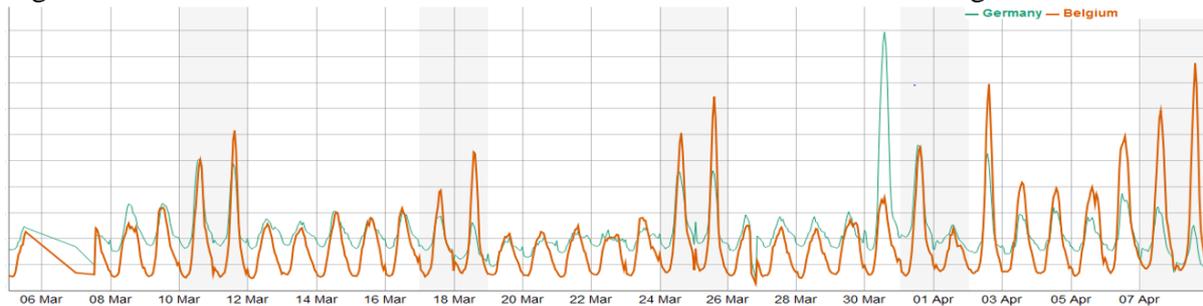

Figure 5.2. Pattern of number of devices in time of a Dutch town at the German border
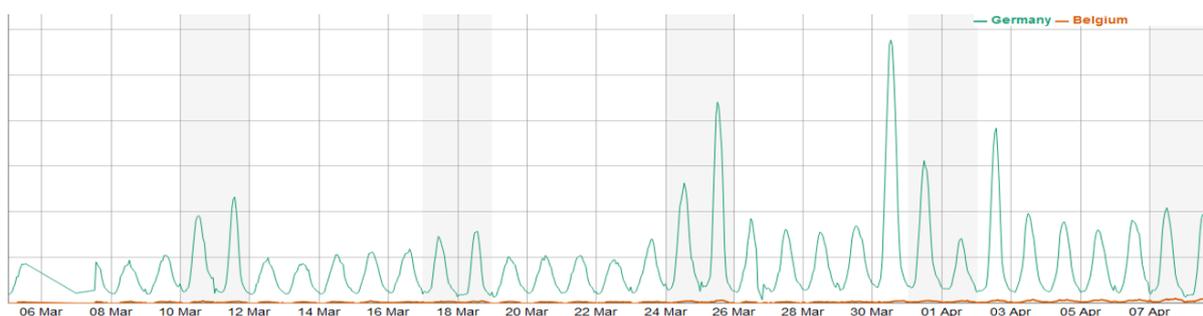
The patterns in time and space are obviously the result of cross-border movements related to different types of activities: tourism, day-trips, shopping, studying or travel to work. Within the project we were not yet able to distinguish between those types of activities. This requires further research and development. But it is feasible without too much effort. One can check how long device stay in a certain location and if this is repetitive. This allows identifying the different activities that are corresponding with these patterns.

Because of the low level of maturity of the methods applied during this project and the provisional status of the data that was used Statistics Netherlands is not able to publish the results of our analysis or present more details about the methods at this stage. Statistics Netherlands will continue to work on this methodology and producing results in the context of another project for DG regio 'City data from LFS and Big Data'. This project focusses on the possibilities to use mobile phone data to measure home-to-work movements and information on Functional Urban Areas. This project includes the statistical offices of Germany and France that are work with mobile phone data This project will publish their findings in January 2019. At that time we will also present results and report more details on the methods used for the cross-border analysis.

4.3. Further work and potential for cross-border information

One of the important elements that has to be carried out is to distinguish foreign devices by type of trip. This could be a day trip for shopping, daily commuting, longer touristic stay or living/studying in the country. This could be done by identifying the weekly pattern and duration of the stay at a certain place. The raw data contains this information. An algorithm has to be developed to make this distinction. This is not straightforward. It needs some testing and requires probably to first develop a few alternatives and subsequently test which method provides the best results.

Another issue how to produce estimates of number of persons rather than mobile phone devices. For the foreign devices this is not easy. There is no easy benchmark for the population. One could think of using the shares of a particular (set of) provider(s) at the home country. To what extent this gives satisfactory results has to be tested.

A further issue is that with national data it is possible to track down where a person entered the country but not where the person came from at the other side of the border. For cross-border data one would like to have both location of origin as location of destination. The correct way of capturing this would be to link mobile phone records at both sides of the border. This is no feasible option. The maximal available data would be to have information on how many travel within the country to the border crossing and how many entered there. This information can then be combined to try to produce origin-destination tables that with most likelihood of begin correct. Such a method can be tested for those countries that this kind of information available for cross-border workers from administrative sources.

4.4. Conclusions regarding mobile phone data

Mobile phone data has a lot of potential in the context of cross-border information. In this field measuring a person movements across borders is a key element to capture. Contrary to traditional data sources that generally measure static situations, mobile phone data captures movements. In addition it allows for a standard methodology that can be used in principle worldwide. This would make it possible to produce international comparable information in an effective way. For this reason that kind of data is extremely powerful to produce cross-border information. We therefore recommend to give the production of cross-border information priority when developing statistics based on mobile phone data.

Our project gave further evidence that is seems possible to develop a universal method to measure cross-border movements with high geographical granularity on real time basis. The movements of foreign mobile phones can be captured into the country. It is in principle possible to distinguish cross-border workers, day-trippers and travellers from each other. This data source can therefore provide consistent information on all these important categories in the context of border regions.

However, before we are able to produce statistics, we still have to solve a number of issues. First of all, the methodology to produce the information has to be developed further. This includes perfecting the measurement of location, distinguishing between several types of cross-border activities, following devices across the border and developing methods to make estimate number of persons rather than devices. Secondly, to get access of the data is not straightforward. This has to be organised as well.

In this project we used aggregated data from a cooperation between Statistics Netherlands and one of the important mobile telephone providers. Within the time frame and restriction of resources available we developed a methodology to produce preliminary results confirming that the source has high potential. The project also proved that to develop such a methodology is a lot of work. We are still at the beginning of the research regarding mobile phone data and a lot has to be done in future.

We recommend to organise a process coordinated and structural development of methods to make use of mobile phone data. NSI's can play an important role in this process to produce official statistics. We advise to take note of the work done in Work package 5 of the ESS-net project on Big data and follow up on the recommendation put forward in their report.

5. CONCLUSIONS AND RECOMMENDATIONS

This pilot project shows what an international collaboration between statistical institutes can achieve in providing evidence in the context of border regions. Important points of policy attention is about increasing cohesion and growth in those regions. More cohesion in border regions will improve quality of life in these generally vulnerable regions. Strong cooperation between actors on both sides of the border improves the possibilities to use the potential in those regions which in turn will lead to stronger economic growth. We decided to focus on the labour market as an area to provide evidence for. That area relates both to cohesion and economic growth.

We investigated the potential of three types of data to provide labour market information for border regions: the labour force survey, administrative data and mobile phone data. The first source is available but could maybe be better explored. The second is available at national level but to what extent can we use it in an international setting? And finally, we use the project to start developing the use of mobile phone data and show what potential such a source could have.

The labour force survey is an excellent source for harmonised data at national level but it does not allow the required geographical granularity to provide valuable evidence in the context of border regions. The limited sample sizes of the survey prevent the production of reliable NUTS 3 data for many regions in the larger countries. This is unfortunate because large countries have many borders. Alternatively, one can use NUTS 3 regions as building blocks to form larger border regions. In the project we have tested this approach and it seems a promising method delivering valuable information. We recommend to carry out further investigations to validate this method in the case of the LFS to prove its usefulness.

Administrative data is used extensively at national level. But what is possible at international level? All countries were able to provide data on incoming cross-border workers based social security files. Producing this data was not straightforward since the population of persons not resident in the country is a rather special population for which at national level less information is produced. Therefore additional work had to be carried out to provide this kind of information. But it is possible and with relatively limited efforts. The results are very interesting. We were able to provide information at NUTS 3 level on the share of employees that work in a specific regions and lives in the bordering country. This information enabled us together with data on citizenship to analyse cross-border work in the different border regions. With a simple set of indicators available for all border regions one can see how border regions differ. In many regions cross-border workers are of a 'classical' nature as border country citizens but in several regions a substantial part involves country nationals living across the border while having their job in their home country. We therefore recommend to put effort in organising the collection of this kind of data at a regular basis.

Within the time frame of the project we not able to analyse the data in depth. Also information from important countries were missing to cover a logical transnational scope. We recommend that efforts are made to collect similar data from the missing countries. In particular data from Switzerland, Austria, Czech Republic and Slovakia would be required to make sensible analysis of cross-border work for France, Germany and Italy.

With this information an analysis could be carried out to make a showcase of the potential and value of having this data on a regular basis. This would be input for discussions with stakeholders about the need of having this data and how this can be organised. We suggest that efforts are made to investigate if it is possible to arrange dedicated analysis project resulting in an report on cross-border work in the EU. This could also include the Nordic countries, since they have this data available.

In the project we had to limit the scope of the data collection for administrative data. But we would like to mention a few elements that could be added to make the data set even more valuable. The first issue concerns the regional detail. For the project we used NUTS 3 regions. For some countries NUTS 3 are quite large. It would make sense to apply more detail by using small regional units. Some possibilities are available: the geographical information could be of administrative, geometrical or functional form. One could think about collecting data on LAU 2 level but confidentiality threshold might be a limit in some cases. In more advanced countries where registers that contain coordinates of each local or productive unit are available even grid data could be explored. Functional areas that cover the whole of the country (such as Labour Market Areas/Zone d'emploi/Travel-To-Work-Areas) could be a valid alternative as indicators on employment are particularly meaningful and comparable at this geographical level.

The second issue is about the place of residence for incoming cross-border workers. We know where they are working but not where exactly they are living in the sending country. This kind of information was excluded from the data collection in this project because it was not readily available for most of the countries involved. However, we know that a number of countries can produce this information with some effort. This is the case for Luxemburg, the Nordic countries, the Netherlands and Germany. Also in Poland, such information could be obtained on the basis of border surveys described in Annex III. We suggest to start an initiative to acquire this information for the countries where this is possible. We recommend to pursue these lines of innovation. At a first stage, it would be best that we first develop the high level of regional detail for a few selected volunteering countries or regions and then scale it up.

We also collected administrative data on population of the border regions. But we come to the conclusion that this data has too limited added value to the regular population data. Furthermore, data on foreigners cannot for all countries be related to the total population resulting in difficulties interpreting the results. Therefore we would not recommend to put effort in collecting this data on a regular basis.

Mobile phone data has a lot of potential in the context of cross-border information. In this field measuring persons crossing borders is a key element to capture. Contrary to traditional data sources that generally measure static situations, mobile phone data captures movements. In addition, it allows for a standard methodology that can be used in principle worldwide. This would make it possible to produce international comparable information in an effective way. For that reason this kind of data is extremely powerful to produce cross-border information. We therefore recommend to give the production of cross-border information priority when developing statistics based on mobile phone data.

In this project we used aggregated data from a cooperation between Statistics Netherlands and one of the important mobile telephone providers. Within the time frame

and restriction of resources available we developed a methodology to produce only very preliminary results. They confirm that the type data source has high potential. It enables the measurement of movements of persons. The movements of foreign mobile phones can be tracked into the country. It is possible to distinguish cross-border workers, day-trippers and tourists from each other. This data source can provide consistent information on all these important categories in the context of border regions. This information can be produced internationally comparable, with high frequency, almost real-time and with maximum geographical detail.

A lot has to be done before we can produce official statistics with mobile phone data. We still have to solve a number of issues. First of all, the methodology to produce the information has to be developed further. Secondly, to get access of the data is not straightforward. We are still at the beginning of the research regarding mobile phone data and much has still to be done. We recommend to organise a process coordinated and structural development of methods to exploit mobile phone data. NSI's can play an important role in this process to produce official statistics. We advise to take note of the work done in Work package 5 of the ESS-net project on Big data and follow up on the recommendation put forward in their report.

One can also think about other Big Data sources to provide cross-border information. This involves for instance traffic loop sensors to measure cross-border traffic. Or web scraping methods to collect tourism information. Poland has concrete plans to explore these methods. In addition, Poland suggested questions that could be added to existing surveys to collect cross-border activities (annex IV).

This project should be seen as a first step in a process to produce cross-border data on a structural basis. It is a proof of concept: a show case of what is possible. The data used and produced in the report does not have the status of a full-grown, mature product. A lot of work still has to be done to reach an acceptable level of completeness both in coverage of border regions and coverage of topics. Also the quality of the data and level of detail of the information presented in the report can and should be improved in a next phase of the process.

Apart from lessons learned on the technical side we have also learned about the organisational side. This project shows that a collaboration of statistical institutes is able to produce new information that is directly relevant for local, national and international policy in cost-effective way. It seems to be an approach that can be taken to develop more cross-border information. In addition to the development of statistics one has to think about how to implement the results. Developing a new product is one thing, putting it into production is another. We advise to make use of the current statistical infrastructure of the European Statistical System and investigate how the production of cross-border data can be best integrated within this system.

ANNEX I : METHOD OF DISAGGREGATION LFS DATA ON NUTS 3 LEVEL FOR POLAND

Method

Disaggregation can refer to temporal as well spatial disaggregation. Further, we present a case of temporal disaggregation but the method cover any aforementioned kind of disaggregation.

Assume that $Y_{l\times1}$ is a vector of low frequency data and $X_{h\times k}$ is a matrix of k variables of high frequency data. Let $A_{l\times h}$ be an aggregation matrix such that AX is low frequency data. We want to estimate high frequency vector $y_{h\times1}$ such that linear constraints of the form $Y = Ay$ are met. For example, assume that we want to disaggregate yearly flow data to quarterly flow data for e.g. GDP. Then $h = 4l$ and we can set $A_{l\times h} = \mathrm{I}_l \otimes 1_4{'}$ where $\otimes$ and $'$ denotes Kronecker product and transposition, respectively. For stack data for e.g. unemployment rate, we would set $A_{l\times h} = \mathrm{I}_l \otimes [0 \quad 0 \quad 0 \quad 1]$.

Usually, the first stage of temporal disaggregation is OLS or GLS regression of AX on Y in a model

$$Y = AX\beta + Ae$$

with respect variance-covariance matrix E of e. In a result, we obtain estimate $\hat{\beta}$ of β of the form

$$\hat{\beta}(E) = [X'A'(AEA')^{-1}AX]^{-1}X'A'(AEA')^{-1}Y.$$

The second stage of temporal disaggregation is benchmarking. It is a solution y of minimizing

$$(y - X\hat{\beta})'E^{-1}(y - X\hat{\beta}),$$

with respect to linear constraint $Y = Ay$. We obtain

$$y = X\hat{\beta} + EA'(AEA')^{-1}(Y - AX\hat{\beta}).$$

Clearly, estimated high frequency error term e depends on previous estimation of β. Poor results of regression may lead to unacceptable benchmarking outcome. Methods of disaggregation are usually based on this two-stage procedure. There is no possibility to set a trade-off between regression an benchmarking.

We propose a method that allows to estimate β and e at the same time and to control where we put an emphasis – on regression or benchmarking.

Consider a model of the form:

$$(1) \quad y = X\beta + e,$$
$$(2) \quad Y = AX\beta + Ae,$$

where y, β and e are unknown. Let B and E be a variance-covariance matrix of β and e, respectively.

We are looking for estimates of β and e by minimizing

$$e'E^{-1}e + \beta'B^{-1}\beta$$

with respect to linear constraint $Z'Ay$. The term $\beta'B^{-1}\beta$ is a regularization term. In a case $B = I$, this term is equal to $\|\beta\|_2^2$ - squared Euclidean norm of vector β. Regularization term is often used to prevent overfitting. E enables to change focus from regression to benchmarking. For $E = \lambda I$, for different $\lambda > 0$ we can set the trade-off between regression and benchmarking. Using Lagrange multipliers method we obtain

$$(3) \quad \hat{e} = EA'(AEA' + AXBX'A')^{-1}Y,$$
$$(4) \quad \hat{\beta} = BX'A'(AEA' + AXBX'A')^{-1}Y.$$

In consequence

$$y = X\hat{\beta} + \hat{e} = (XBX'A' + EA')(AEA' + AXBX'A')^{-1}Y.$$

What is worth to notice, (3) and (4) show that estimates of β and e depend on a common component $(AEA' + AXBX'A')^{-1}$ which defines the trade-off between regression and benchmarking. Checking the constraint (2) we have

$$Ay = (AXBX'A' + AEA')(AEA' + AXBX'A')^{-1}Y = Y.$$

Model

Our first aim was to disaggregate full-time and part-time workers from NUTS 2 level to NUTS 4 level using auxiliary data from Social Insurance Institution of Poland and the size of population with respect to gender with a use of model described previously. Basing on estimates from the model we used auxiliary data from Social Insurance Institution of Poland again to produce additional statistics on the number of part-time workers with respect to working hours and the structure of employed with respect to the number and type of contract.

Working hours were divided into following groups:
1. Less than 1/8-time or 1/8-time employed
2. Over 1/8-time but less than 1/4-time employed
3. 1/4-time employed
4. Over 1/4-time but less than half-time employed
5. Half-time employed
6. Over half-time employed
7. Not assigned (employer is obliged to assign working hours during employee registration into the system)

Contracts were divided into following groups:
1. Labour contract
2. Task-specific contract
3. Other civil law contract

The data is presented with respect to one and two contracts. Share of people with three and more contracts is less than 8% for whole country and covers 27 subgroups. Thus in a general structure each subgroup covers very small share of population. Hence it was omitted.

ANNEX II: LABOUR MARKET AREAS FOR ITALY

A SPATIAL UNIT FOR REPORTING ON EMPLOYMENT AND UNEMPLOYMENT: THE EXAMPLE OF ITALY

Labour market areas (LMAs) are sub-regional geographical areas where the bulk of the labour force lives and works, and where establishments can find the largest amount of the labour force necessary to occupy the offered jobs. They respond to the need for meaningfully comparable sub-regional labour market areas for the reporting and analysis of analysis. LMAs are defined on a functional basis, the key criterion being the proportion of commuters who cross the LMA boundary on their way to work. A Labour Market Area is a functional geographic area or region beyond the administrative boundaries defined for purposes of compiling, reporting and evaluating employment, unemployment, workforce availability and related topics. It is an economically integrated spatial unit within which residents can find jobs within a reasonable commuting distance or can change their employment without changing their place of residence[3].

They are functional regions that stem from the aggregation of elementary geographical units (municipalities, LAU) on the basis of their level of spatial interaction measured by commuting to work flows through quantitative methods. The guiding idea is to maximise the flow inside the area (internal cohesion) and minimise it outside (external separation) according to a predefined rule.

The needs for drawing LMAs may range from the simple description of the territorial objects to more complex and policy oriented disseminations. LMAs have long being recognized as relevant for assessing the effectiveness of local policy decisions. Possible objectives when drawing LMAs may range from:
- Identify the territorial boundaries of labour market for planning and analysis purposes;
- Define an economic geography to support policy making.

LMAs are developed through an allocation process based on the analysis of commuting patterns. 2011 LMAs are based on commuting data stemming from the 15th Population Census using a new allocation process, an evolution of the previous algorithm: they are 611 distinct areas. The implementation of the same allocation process on 2001 commuting data gives rise to 683 zones; once again a reduction in the number of areas is observed, and it is due to the increase of the trend in more and longer distance commuting.

Istat has been releasing the geography of LMAs since the '80s to investigate employment at local level. Given that LMAs are not a planned estimate domain as they cut across survey strata, small area estimation approach that gains strength thanks to neighboring areas as been adopted (D'Alò et al. 2017). The estimates preserve the spatial distribution of target variable. The time series of employment and unemployment rate estimates for the whole Italian territory at the level of LMAs is regularly updated from 2006. The time series 2006-2017 is available at: https://www.istat.it/it/archivio/217437.

As an example Figure A1 contains the estimates of unemployment rate al LMA level for year 2016, based on data from the Italian Labour Force Survey.

Figure A1 Estimates of unemployment rate al LMA level for year 2016, based on data

---

[3] Cros Portal: https://ec.europa.eu/eurostat/cros/content/labour-market-areas_en

from the Italian Labour Force Survey. The grey lines represents the border of the LMAs, in black the NUTS3 ones.
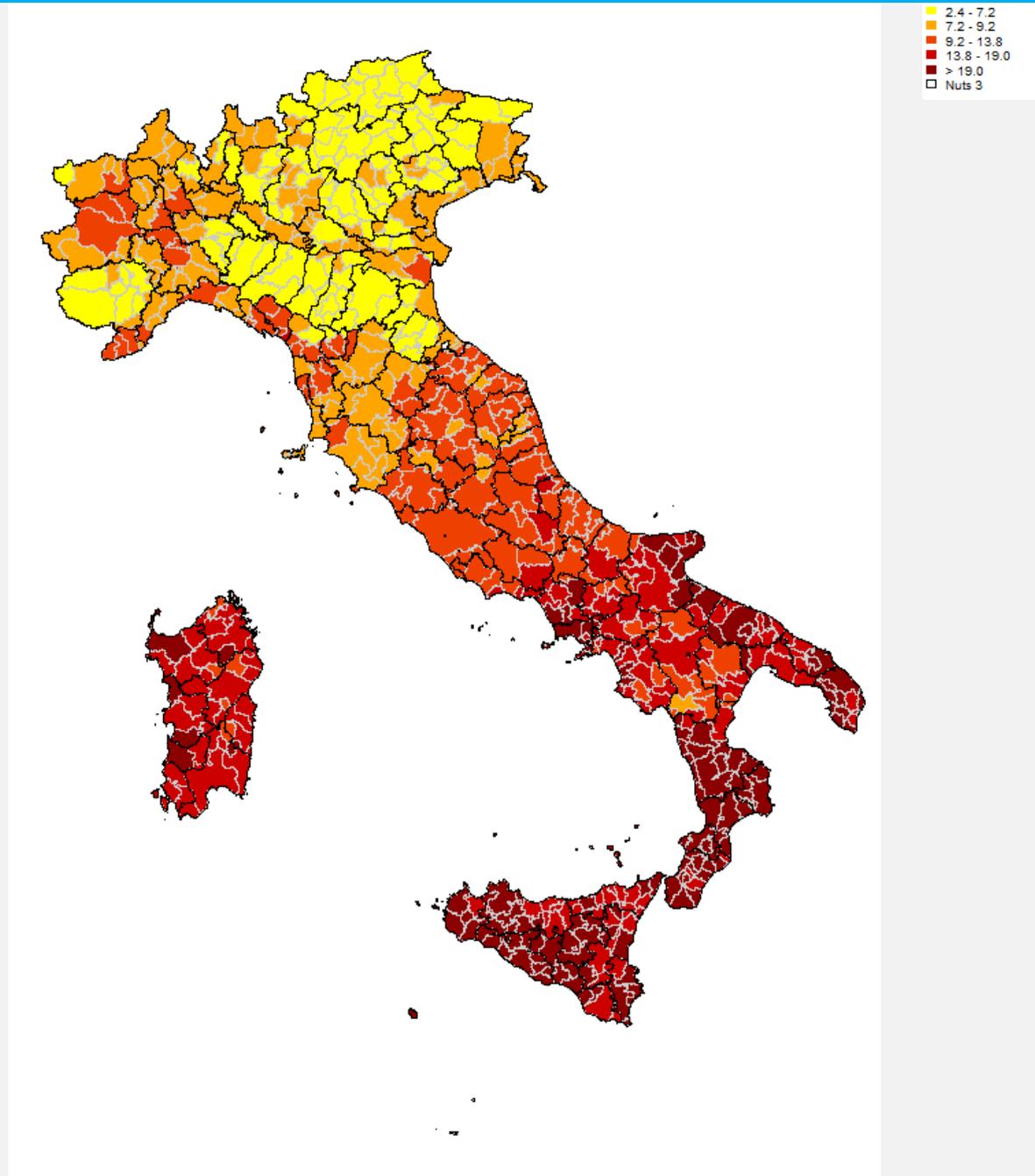
As part of the Grant "EU-TTWA method: improvements, documentation and sharing knowledge activities" funded by Eurostat and awarded to Istat the Italian National Statistical Institute, a freely available R package LabourMarketAreas for the delineation of the areas has been developed; more information on the project is available in Istat web page (http://www.istat.it/en/archive/182743). The grant is part of a system of grants in the EU with the aim of supporting the ongoing development of a methodology for the creation of LMA and test the application of methods nationally in the participating National Statistical Institutes. The goal is to arrive at a harmonised EU-wide definition for

labour market areas; guidelines for the delineation of LMAs have also been produced. More info at: https://ec.europa.eu/eurostat/cros/content/labour-market-areas_en

More general information on LMAs and example of indicators produced at this geographical level can be found in: Franconi, L., Ichim, D. and D'Alo', M. (2017). Labour Market Areas for territorial policies: tools for a European approach. Statistical Journal of the IAOS, Vol. 33, No. 3, pp. 585-591. https://content.iospress.com/articles/statistical-journal-of-the-iaos/sji160343

A general presentation of the LMA concept has been prepared for the Group of Experts on Population and Housing Censuses (26 - 28 September 2018); in this paper a brief example of how to use the R package LabourMarketAreas for the delineation of the areas is sketched.

References

D'Aló M, Falorsi S, Solari F. Space-time unit level EBLUP for large data sets. Journal of Official Statistics 2017; 33(1): 61–77.

ANNEX II: INTEGRATED SURVEYS ON TRIPS IN POLAND

Integrated surveys on trips cover the surveys: "Trips made by non-residents to Poland. The movement of vehicles and persons at the Polish border with the countries of the European Union" and "Participation of Polish citizens (residents) in trips".

The survey on the participation of Polish citizens in trips is carried out in households and the survey of foreigners coming to Poland (survey at the external and internal borders of the EU's on the territory of Poland) is carried out in the vicinity of selected border crossings. At the same time, registration of border crossings is carried out at the borders with the countries of the Schengen zone. The surveys are conducted by Statistics Poland and the Statistical office in Rzeszow in cooperation with the National Bank of Poland and the Ministry of Sport and Tourism.

The primary objective of the surveys is to provide methodologically consistent information adjusted to diverse needs of users in the field of domestic and outbound trips of Poles and foreigners coming to Poland. The collected data are used for the needs of tourism, national accounts, balance of payments and transborder areas statistics. The information obtained in the surveys is used by government authorities at the central and regional level, and by local government in the border regions to formulate the development strategies.

SURVEY IN HOUSEHOLDS ON PARTICIPATION OF POLES IN TRIPS

The aim of the survey on the participation of Polish citizens in trips is to determine the scale of participation of residents in trips, the characteristics of domestic trips and trips abroad taking into account expenses related to trips.

The survey is conducted in households, on a quarterly basis.

The members of the one-person or multi-person household in randomly selected dwellings are the unit of observation in the survey.

Objective and scope of the survey

The survey examines the participation of Polish citizen in domestic and outbound trips. The survey covers trips that last less than a year, with the exception of outbound trips related to education or treatment and covers trips with at least 1 overnight stay and same-day trips (without an overnight stay). Domestic same- day visits are surveyed cyclically every 3 years.

The questionnaire "PKZ" is used for the survey.

Among other things, the interview questionnaire includes questions concerning:
- socio-demographic characteristics of household members,
- completed domestic and **outbound trips** as well as their characteristics, i.e.:
    - direction (domestic/outbound),
    - destination,
    - duration of stay,
    - **main purpose, of which seasonal or border work,**
    - type of accommodation used,
    - main means of transport,
    - how the trip was organized,
    - expenses incurred by type, and in the case of outbound trips expenses incurred on services provided abroad

Survey method. Sample design

The survey on the participation of Poles in trips is a sample questionnaire survey, carried out by means of face-to-face interview by interviewers in the whole country. Participation in the survey is voluntary.

A quarter is the reference period and the survey is carried out in the month following the quarter.

The survey is carried out on a sample of 18 750 dwellings, which are drawn on a quarterly basis. The sample is drawn from a frame built on the basis of a set of Census enumeration areas (from which Census enumeration areas with zero flats were excluded), with a two-stage sampling with stratification on the first stage. Census enumeration areas, or sets of

Census enumeration areas, with a minimum of 30 dwellings are the first-stage sampling units (primary sampling units – PSU). Census enumeration areas which do not fulfil this condition are combined into a set within the same statistical division. The second-stage sampling units are dwellings.

Census enumeration areas are sorted by strata. Strata containing border areas are divided into two parts: border part and internal part. A border part consists of gminas located not further than 30 kilometres from the border or the coastline. Areas in a coastal zone without access to marine connection with foreign countries are treated as an internal part. If a part of gmina is situated in a distance between 30 and 50 kilometres from the border line, it is included in the border part as well. This zone has been mapped out along Polish border based on the definition of the border area from the Regulation of the European Parliament and of the Council of 20th December 2006 and the results of the survey of goods and services turnover in border traffic at the EU's external border in Poland.

In order to obtain a sufficient number of questionnaires for same-day trips abroad, the sample is doubled in border zones as these areas see the highest number of same-day trips. Therefore, half of the sample is allocated to the border strata and the other half to the central strata. Within each of these strata the sample allocation is proportional to the number of dwellings in a stratum.

Sample selection

Poland	Border area

dwellings selected to the survey
- 5 - 15
- 16 - 25
- 26 - 35
- 36 - 45
- over 45
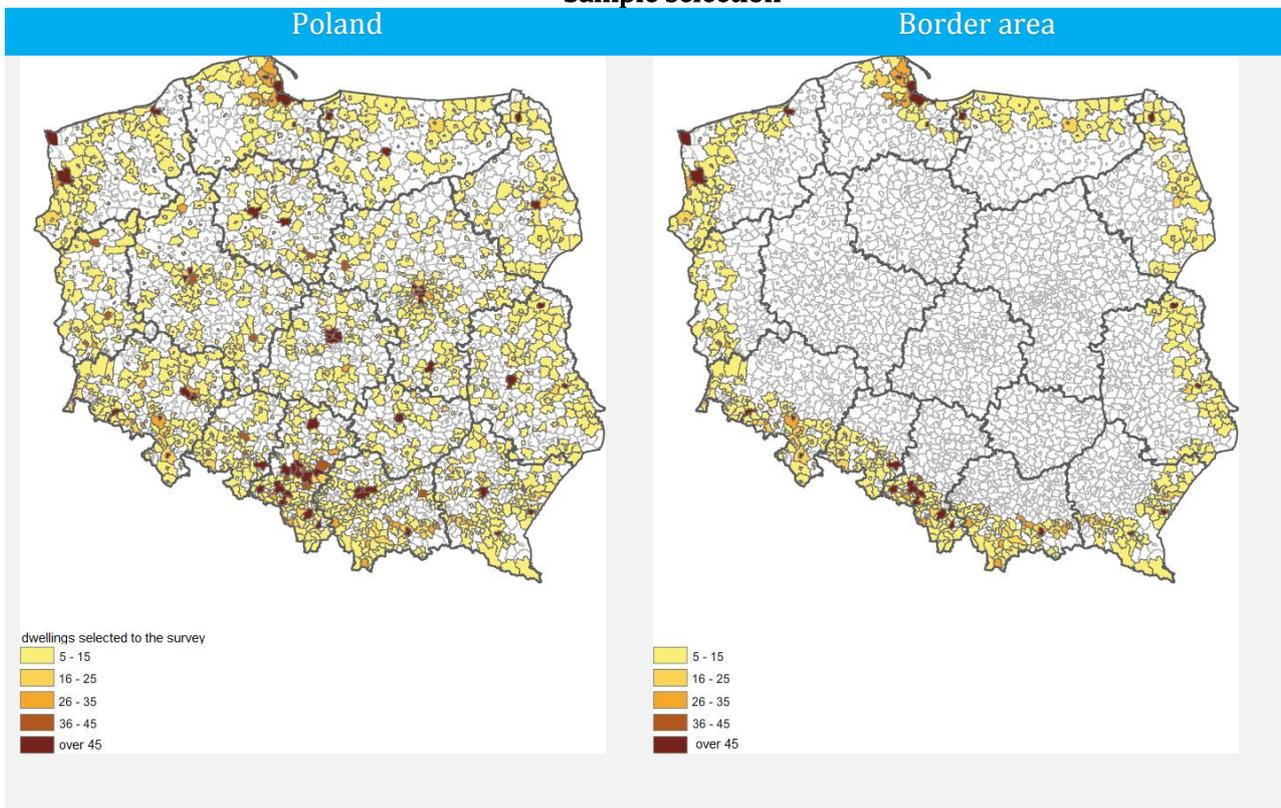
Due to the requirement to obtain a sufficient number of questionnaires with information on foreign trips, there was a need to use the random route method in the survey. According to this technique, if in the randomly selected dwelling (the so-called starting point) the members of the household did not travel, the interviewer goes to the next dwelling in order to conduct the interview, but it is always guided by a well-prepared algorithm for the selection of subsequent dwelling for the survey. The interviewer visits a maximum of 5 dwellings. The survey do not provide for a reserve sample.

SURVEY ON THE BORDERS ON BORDER TRAFFIC OF VEHICLES AND PERSONS AS WELL AS TRIPS MADE BY FOREIGNERS TO POLAND

The aim of the survey is to estimate, on a quarterly basis, the number of trips made by non-residents to Poland, as well as the expenses incurred in connection with the trip to Poland, the characteristics of the stay of visitors from abroad, to obtain information about persons crossing the border as part of local border traffic, and to estimate the border traffic at the Polish border with other countries in the European Union

The survey of border traffic of vehicles and persons
The border traffic survey covers persons and vehicles crossing the Polish border with the countries of the European Union at selected border crossings.
The following are recorded:
- the number of vehicles crossing the Polish land border in both directions (arrival and departure),
- the symbol of the country of the vehicle registration,
- the number of persons traveling in vehicles,
- pedestrian traffic in both directions – registration of persons, including the question about the country of residence.

For the survey on the internal borders of the EU's on the territory of Poland questionnaires BRG and BRGp are used.

The survey of trips made by foreigners (non-residents) to Poland
The survey of trips made by foreigners to Poland is a sample and questionnaire survey. It is carried out in the vicinity of selected border crossings on the EU's internal and external border in Poland including seaports and airports, with the use of direct interviews conducted by interviewers. Participation in the survey is voluntary.
The survey includes foreigners (non-residents) leaving Poland through the EU's internal and external border on the territory of Poland:
- tourists
- same-day visitors,
- **seasonal and border workers**,
- business travellers as members of the crew of a means of transport (train, ship, plane).

The survey covers persons crossing the border as part of the local border traffic and the ones possessing the Card of the Pole. The survey does not include people residing in Poland for over one year (with the exception of students and people traveling for health reasons).
In the survey the questionnaire PDP - Trip to Poland is used, which includes:
- the country of residence;
- characteristics of the persons traveling according to the main purpose (of which border and seasonal work), length of stay, sex and age,
- characteristics of the persons crossing the border as part of the local border traffic and possessing the Card of the Pole;
- type of accommodation and means of transport used;
- how the trip was organized
- expenses connected with the trip to Poland and incurred in the country of permanent residence;
- expenses incurred in Poland for the purchase of services (including accommodation, catering, transport and the other) and goods;
- the distance from the place of purchase and place of residence to the surveyed crossing;
- the frequency of crossing the border.

Due to the fact that the survey is carried out among foreigners, the questionnaire PDP is prepared in 19 language versions.

Survey method, sample selection
The primary source of information for the survey of trips made by foreigners to Poland is the PDP questionnaire "Trip to Poland". In the survey of border traffic on the EU's internal border the forms BRG and BRGp are used.

The estimation of the results is based on secondary results of "Survey of border traffic" and additional information of Border Guard Main Headquarters on border traffic on crossings on the EU's external border in the days when the survey is conducted. To determine the size of border traffic at airports and seaports data on passenger traffic at these ports are used. The estimation of border traffic is also made based on data on the use of tourist accommodation establishments from the survey "Tourist accommodation establishments and their occupancy" and on Big data.

Sample selection
The survey of the crossings of the border is a sample survey, which is carried out at selected border crossings on the EU's internal land border in Poland. It is based on registering (counting) of crossings of the border both by people and vehicles (in both directions). The survey is conducted in each quarter.

Due to the lack of a frame population, the survey of trips made by foreigners to Poland is carried out on "elusive" population, however, due to the knowledge about the places where the survey is conducted, adequate representativeness of the sample is ensured.

The survey of trips made by foreigners to Poland is a questionnaire survey, carried out in the vicinity of selected border crossings on the EU's internal and external border in Poland, including seaports and airports, with the use of direct interviews conducted by interviewers. Participation in the survey is voluntary. Persons are surveyed using systematic sampling. In the case of refusal of the selected person to participate in the survey, another person is being surveyed. For particular border crossings sampling intervals were determined taking into account the projected volume of traffic and possibility of the interviewer to conduct the interview at a particular time.

It is important to select a proper sample and to obtain good quality data in order to develop an appropriate method for estimating the border traffic. The survey of this phenomenon must therefore be limited to the necessary group of crossings and certain time periods (days and hours) in a year, but at the same time ensuring the quality of the results.

In order to select border crossings for the observations a typology was carried out. In the first stage the crossings were divided by type of border, selecting three types of crossings:
- land border – road, rail and river crossings,
- maritime border – ports,
- air border – airports.

The land crossings were divided according to the nature (permeability) of the borders and the neighbouring country, and therefore located at:
- the EU's external border on the territory of Poland, including Polish-Russian border, Polish-Belarusian border, Polish-Ukrainian border;
- the internal border of the EU and Poland, including Polish-Lithuanian border, Polish-Slovak border, Polish-Czech border, Polish-German border.

Then, on the basis of data of Border Guard Main Headquarters on border traffic from 2007, characteristics of border crossings was made.

Border land crossings located on internal borders of the EU and Poland whose share of border traffic with the neighbouring country was greater than 1% were the subject of further analysis. Out of the 55 crossing points meeting this condition, 54 are road crossings and 1 railway.

To obtain the appropriate quality of estimates, the rotation of border crossings was introduced. It consists in the fact that among the randomly selected border crossings for the given border section there are also border crossings that were surveyed in the previous year. This approach ensures the continuity of information about the changing conditions prevailing at the crossings. In the vicinity of randomly selected border crossings, interviews of foreigners leaving Poland are also carried out.

Road crossings at the EU external border in Poland are drawn based on the typology and current **KGSG** data on the movement of people and vehicles at these crossings. As on the external border, the Border Guard still keeps the registration of persons crossing the border, only surveys of foreigners regarding their trip to Poland are carried out.

Due to its specificity, the survey of border traffic at airports and sea ports was prepared separately using appropriate data sources (in particular reporting on air and maritime

transport). When selecting border crossings, the possibility of obtaining relevant data is taken into account in the structure of traveling people.

Due to the small share of railway crossings in border traffic, no permanent survey at these crossings is carried out.

Organization of the survey at the border
Internal (land) border

The border survey (border traffic at the internal land border and questionnaire survey among foreigners) is carried out by 7 randomly drawn days (from Monday to Sunday) in the quarter at selected border crossings. During two days, 24-hour surveys are carried out - one on a weekend day (Saturday / Sunday) and one on a working day. During the 24-hour survey, only the vehicle counts are carried out and people traveling there, as well as people traveling on foot; no survey of foreigners takes place. On other days, the survey is carried out from 6 a.m. to 10 p.m. and includes registration of border crossings and surveys, implemented alternately.

External border, seaports and airports

The survey carried out at the EU external border on the territory of Poland, in seaports and airports includes interviewing foreigners leaving Poland. It is carried out during 7 randomly drawn days in each quarter - from Monday to Sunday. The survey time is determined on the basis of observation of traffic at a given border crossing. The survey is assumed to be conducted at road crossings at the external border for 8 hours, while at airports and seaports, it is adapted to the volume of passenger traffic.

Generalization of the results
Estimation of the border traffic at the internal border

For the generalization of survey results the estimation of data from survey of quarterly periods was used. Annual estimates are the sum of quarterly results.

The generalization process of the survey result is preceded by an analysis of the collected data, the parameters of sample and data processing. At this stage of work estimation and imputation methods are used apart from the formal method.
These methods are used for:
- vehicles with an unidentified symbol of the country of registration. These vehicles are marked with the symbol XX. The estimates of this range, i.e. for Polish and foreign number plates are made on the basis of the structure of vehicles traffic with identified symbols of the country of the registration;
- vehicles with an unidentified number of people. In cases when the interviewer failed to determine the number of persons in a vehicle, it was marked with the symbol X. In such cases, an estimation of the number of persons is based on an average number of people in the same type of vehicles with known number of people for a given day and border crossing. If calculation of such an average is impossible, the next (higher) level of data aggregation is taken into account, i.e. the average number of people in vehicles of the same type for the border crossings during the week, for a given section of the border or the total value for all internal border;
- pedestrians crossing the border. The estimation of the number of Poles and foreigners was based on the structure of pedestrian traffic at selected border crossings;
- interval higher than 1 ($N>1$). In the case when, for example, $N=2$ (every second vehicle was interviewed) the records are doubled. This solution allows us to equally treat each record and its representativeness for calculation is the same.
- estimation of 16-hour survey into 24-hour periods.

Estimation of border traffic volume at the EU's internal border on the territory of Poland at land border
crossings

The results of counting and estimation of data for each crossing are summarized and analysed according to traffic intensity and location on a given section of the border with a

neighbouring country. Additionally, data from Big data sources are used. For estimating total border traffic of people, the regression analysis is used. The purpose of regression analysis is to examine a relation between the total number of crossings at a given section of the border and the number of crossings at the selected border

crossings. As data of the Border Guard Main Headquarters before Poland's entry into the Schengen area are the base, it is necessary to include additional information in the analysis. This approach is justified by the fact that the structure of the traffic and their share in total traffic could have changed significantly on individual border crossings.

Considering that Poland borders with 4 EU countries (i.e. Germany, Czech Republic, Slovakia and Lithuania) and the traffic at the sections of the border with these countries is very diverse, the total road border traffic at the EU's internal border EU on the territory of Poland is the sum of the values for individual sections of the border. Data on border traffic of Poles are estimated on the basis of information from the survey conducted in households and counting of vehicles and people on selected border crossings. The border traffic of foreigners leaving Poland, however, is estimated using regression analysis for each of the section of the border and on the basis of data from counting of vehicles and people, as well as other available data sources (of which Big data).

The estimates of the number of foreigners by country are based on the analysis of identified countries, i.e. the registration numbers of vehicles and pedestrians' declarations of country of origin counted on selected border crossings. In addition, information about country of origin is combined with the database on the use of collective accommodation establishments and the structure of people who responded to the questionnaires regarding the made by foreigners to Poland. A synthetic summary of the information serves as the basis to determine traffic intensity of people of particular countries.

Estimation of traffic intensity at seaports and airports

Due to the specificity of seaports and airports it was necessary to apply an individual approach to estimating traffic intensity at these border crossings.

In the case of seaports, data concerning traffic intensity are received from reporting of Statistics Poland and the Border Guard Main Headquarters (BGMH). Data on passenger traffic in seaports are compiled by port, place of departure or arrival and directions. The passenger structure is, however, obtained from the survey conducted in selected ports, data of BGMH and operators serving maritime traffic.

Data for air traffic are a comprehensive compilation of information from various sources, which include:
- information from airports on passenger traffic (the number of passengers handled in regular and charter traffic),
- data from the Civil Aviation Authority on passenger traffic (information on domestic and international traffic, flight directions and cities of destination),
- information from the Border Guard about traffic intensity at border crossings and selected countries,
- structure of passengers obtained from the survey conducted at selected airports.

Generalization of the survey results for internal and external borders

The results of the survey of the intensity of border traffic of vehicles and people and trips made by foreigners to Poland, as well as expenses incurred by them on these trips are compiled for quarterly and annual periods.

The basis for estimating the results on external land border is data obtained from the questionnaires and information from the Border Guard Main Headquarters on border traffic. On the internal land border, the estimated results are based on data obtained from questionnaires and data on traffic of vehicles and people counting, which are then generalized to respective border crossings and sections of the borders. These data include the number of foreigners crossing the border by border crossing, direction and types of traffic (way of crossing the border) in the surveyed quarter and days in which the questionnaire survey was carried out. Data are generalized separately for each stratum.

The method of estimating trips made by foreigners to Poland and their expenses to all countries of the world

In order to estimate the results of trips made by foreigners (non-residents) to Poland and their expenses, additional data on the use of tourist accommodation facilities, data of the Border Guard Headquarters on foreigners crossing the border by country of origin, data on crossings of the internal EU border in Poland, obtained from the survey of border traffic, as well as information from airports and the seaports are taken into account. The data sources listed above contain information about trips made by foreigners to Poland from over 190 countries of the world.

In the first step all countries in the world were divided into 19 categories according to the different specificity of average expenses, type and length of stay, purpose of the visit, distance from Poland, etc. Among these categories, the neighbouring countries of Poland were distinguished separately. The remaining European countries were divided into 4 groups (Eastern Europe, Southern Europe, Western Europe and Northern Europe), Africa - into two groups (North Africa and South Africa), Asia - into two groups (Middle East Asia and Far East Asia), America into 3 groups (North America, Central America and South America) and Australia and Oceania were also separated. In some cases, calculations are made on combined categories due to the specificity of the topic (e.g. calculation of average expenses for same-day visitors).

In the next step, the number of trips for each country is compared based on the data of the Border Guard Main Headquarters. These values, in turn, are adjusted based on the KT-1 report – a report on the use of tourist accommodation facility, containing information on the number of foreigners using accommodation facilities.

Then, the number of overnight stays of tourists from particular countries is calculated. The principle of calculating the average length of stay (number of nights per trip) to particular groups of countries was adopted.

The total expenditure per country is the product of average expenses for the category in which the country is located and the estimated number of trips for this country.

Based on the results of the survey "Trips made by non-residents to Poland. The movement of vehicles and persons at the Polish border with the countries of the European Union" and "Participation of Polish citizens (residents) in trips" reports are developed in the form of quarterly information and annual publications. The studies contain the results of the survey of people crossing the external and internal border of the European Union in our country. In particular, they contain information on characteristics of expenses incurred in Poland by foreigners (non-residents) and Polish citizens (residents) abroad, as well as border traffic, including the purpose of trips or the frequency of crossing the border.

Publications are available at:
1. http://rzeszow.stat.gov.pl/publikacje-i-foldery/obszary-przygraniczne/ruch-graniczny-oraz-wydatki-cudzoziemcow-w-polsce-i-polakow-za-granica-w-2016-r-,15,3.html
2. http://rzeszow.stat.gov.pl/opracowania-biezace/opracowania-sygnalne/obszary-przygraniczne/ruch-graniczny-oraz-wydatki-cudzoziemcow-w-polsce-i-polakow-za-granica-w-iv-kwartale-2017-roku,14,16.html

ANNEX IV: POLISH PROPOSALS ON ALTERNATIVE SOURCES

Big Data sources concerning cross-border areas in Statistics Poland

1. Use of sensors of road traffic to estimate cross-border traffic on the Polish-German and Polish Lithuanian border

2. The use of the Web scraping method to create a database of accommodation offers in selected countries

Possibilities to add questions on cross-border issues in surveys

Cross-border trade in agriculture

Survey	Question in the questionnaire	Solution
The European Union Statistics on Income and Living Conditions (EU-SILC)	Did your homestead get income?	response Yes should have submenu YES - from cross-border areas
Household Budget Surveys (HBSs)	In the section on agriculture	Add a column: country of purchase including cross-border areas (selection from the country dictionary)
International trade in services (ITS)	Services provided for agriculture, forestry and fishing	Country where we deliver services including cross-border areas
Business tendency survey in agriculture	Does the farm sell the agricultural products	Adding a response: YES - sales to foreign customers including cross-border areas
	For what purpose have been allocated funds in the last 12 months:…..	To the answer 02, 07, 08, 09, 10 add the variant : "... at a foreign counterparty"

Cross-border on-line shopping

Survey	Question in the questionnaire	Solution
ICT usage in households and by individuals (isoc_i)	From which sellers you have bought or ordered online goods and services for private use in the last 12 months	Add a country including cross-border areas
Household Budget Surveys (HBSs)	In the section on cash expenses	Add a column: country of purchase including cross-border areas (selection from the country dictionary)
International trade in services (ITS)	Purchase / sale of computer software and games downloaded from the Internet or on-line	Country where we deliver services including cross-border areas

Labour force survey (LFS) in the section on job search

Survey	Question in the questionnaire	Solution
Labour force survey (LFS) Section: job search	In order to find / take up a job, would you be willing to change your place of residence?	Adding a response: „abroad, within a cross-border area"
	Why you are looking for a job?	Adding a response: „I am looking for better financial conditions in the cross-border area"
	In order to find / take up a job, would you be willing to change your place of residence?	Adding a response: „Abroad, in the cross-border area"

Publications Office