Håvard Hungnes Lien, Eurostat

Do the immigrants take our jobs?

Background, hypothesis for testing, methodology, data sources

A recurrent claim in the public debate is that 1) immigration leads to higher unemployment among nationals, and 2) industries which employ many immigrants see decreases in earnings. This paper seeks to examine to what extent these claims can be substantiated, by checking if there is a monotonic correlation (using Kendall Tau b) between immigration and native unemployment, and between immigration and earnings.

Obviously, correlations are not to be taken as proof of causation.

To this end we will be using labour force survey data for the time period 2006 – 2015 and earnings statistics data for the period 2008 - 2015. The starting points are chosen because 2006 is the first year we cover all member states in the LFS data set, and 2008 is the last NACE revision.

As we will only look at effect on unemployment and on wages, this paper will only consider the economically active population, aged 15-74. Migration status is derived from the variable NATIONAL.

The LFS has limitations in capturing all effects of immigration, as it in most countries is based on the usual resident population. This analysis can therefore not take into account effects of cross-border commuters, short term / posted workers, or persons living in communal households.

Kendall's tau-b determines whether there is a monotonic relationship between two variables (the two variables rise or fall in parallel, but not necessarily linearly). The variables must be at least ordinal scale. It is a nonparametric (making no assumptions about the probability distributions of the variables) alternative to the Pearson's product-moment correlation, for cases when data is not linearly related, has outliers, and is not normally distributed. This is the case for the data in this paper. The Kendall correlation is better than the Spearman correlation for cases with small sample sizes and tied ranks (which also is the case in this paper).[1] Its values go from -1 (perfect negative correlation) to 1 (perfect positive correlation). A correlation value of 0 means no correlation. In practical terms, results in the range -0.4 to 0.4 are considered as weak correlations, and values below -0.6 and above 0.6 are considered as strong correlations.

[1] There is a vast literature on correlation coefficients. See for instance 'Prediction of reliability – the pitfalls of using Pearson's correlation' by Hryniewicz and Karpinski http://www.ein.org.pl/sites/default/files/2014-03-18.pdf for a discussion of Pearson, Spearman, and Kendall.

Main findings

- Only five countries show a high positive correlation (above 0.6) between EU immigration and native unemployment combined with an immigrant population of some note (at least 2% of the active population are EU immigrants)
- Three countries show a high positive correlation between non-EU immigration and native unemployment
- There is not much regional variation inside countries regarding these correlations
- Earnings correlate negatively with EU immigration in two country/NACE groups, and positively in seven
- Earnings correlate negatively with non-EU immigration in five country/NACE groups, and positively in one

This means that the hypothesis on unemployment is not substantiated for a large majority of the countries, and further that there is somewhat more empirical basis for saying that immigration correlates with increased earnings than with decreased earnings.

Immigrant populations in 2015, country level

In order to provide some context to the analysis, we will first take a quick look at the level of immigration in 2015. In map 1 we see that economically active EU immigrants (intra EU mobility for reasons of working) range from the extreme outlier of Luxembourg (47 per cent of the total active population) to practically no EU immigration in Bulgaria, Czech Republic, Estonia, Greece, Croatia, Latvia, Lithuania, Malta, Hungary, Poland, Portugal, Romania, Slovenia, Slovakia, Finland, Macedonia, and Turkey (all less than 2 per cent).

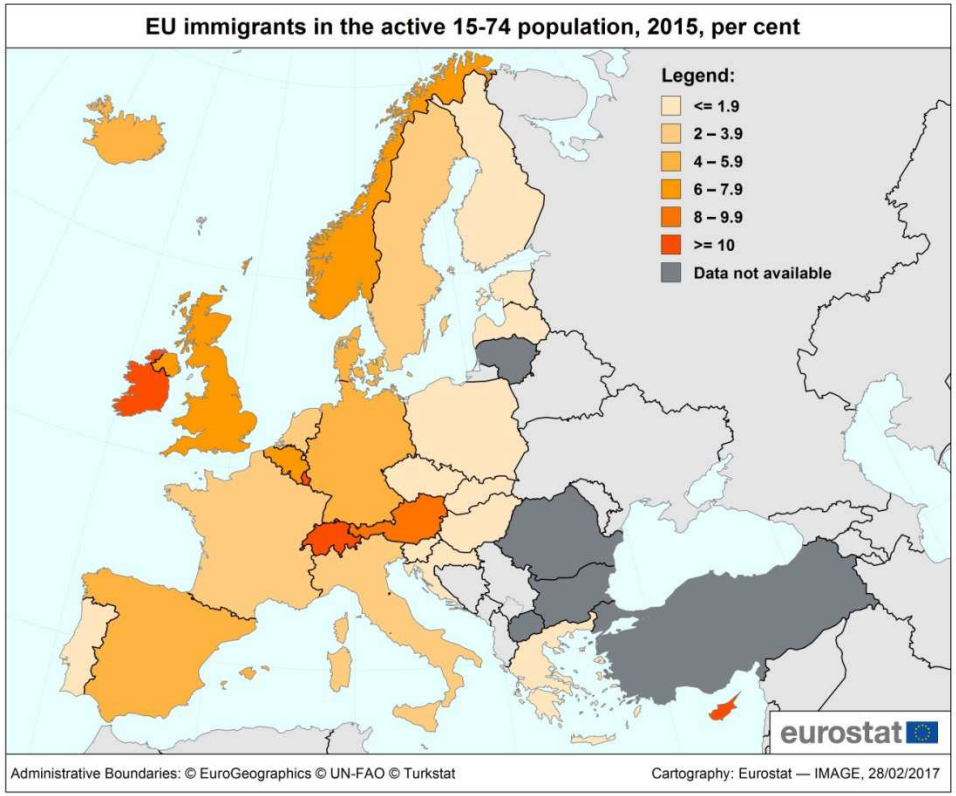
In-between these extreme we have the four intermediary group, in increments of two percentage points. France, the Netherlands, Sweden and Italy are in the 2.0 to 3.9 group. Iceland, Denmark, Spain, and Germany follow in the 4.0 to 5.9 group. Further up we find, in the 6.0 to 7.9 group, United Kingdom, Norway, and Belgium. Finally, Austria is alone in the 8.0 to 9.9 range. Luxembourg is at the top, as said, with 47 per cent.

The overall pattern is that EU immigration for purposes of work goes in the direction of the older member states.

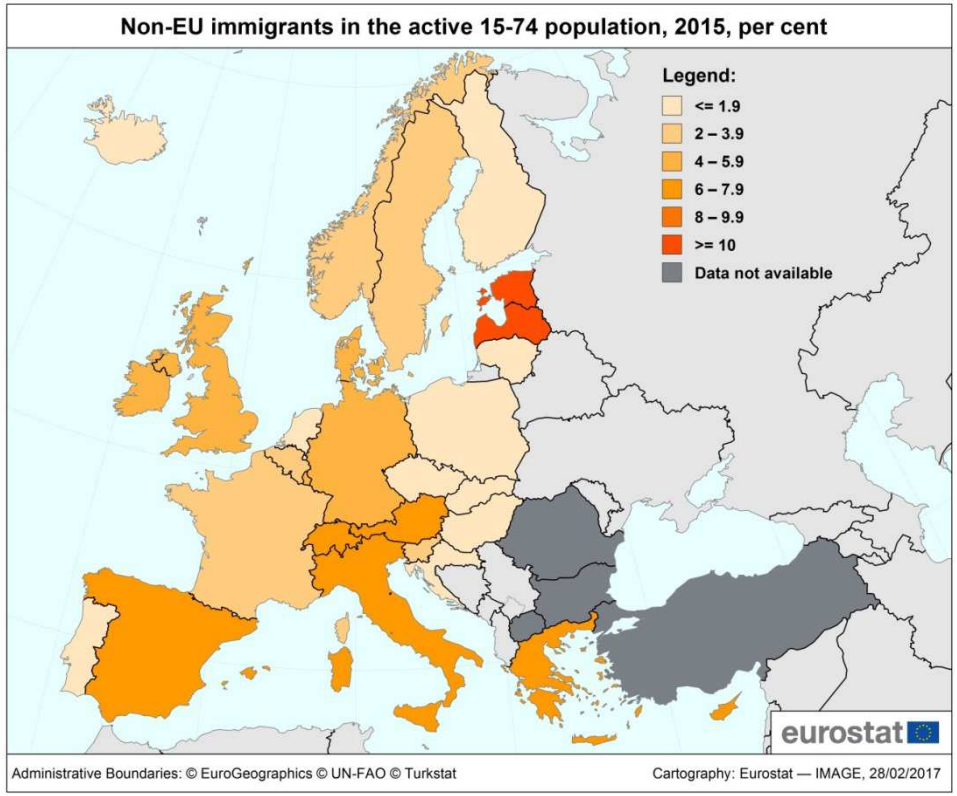
Regarding non-EU immigrants we see in map 2 that the pattern is different. The highest levels are in the Baltics (with the notable exception of Lithuania), followed by a Mediterranean/Alpine group of Cyprus, Spain, Italy, Greece, Switzerland and Austria. At the low end of the scale (less than 0.5 per cent) we find Bulgaria, Croatia, Hungary, Poland Romania, Slovakia, Macedonia, and Turkey.

In conclusion we find that, with the exceptions of the three Baltic countries, Italy, Greece, and Spain there are more EU-immigrants than non-EU immigrants. Overall, we see that there are more immigrants in Western Europe than in Eastern Europe.

Map 1: EU immigrants, 2015, proportion of active 15-74 population, country level, per cent



Map 2: Non-EU immigrants, 2015, proportion of active 15-74 population, country level, per cent

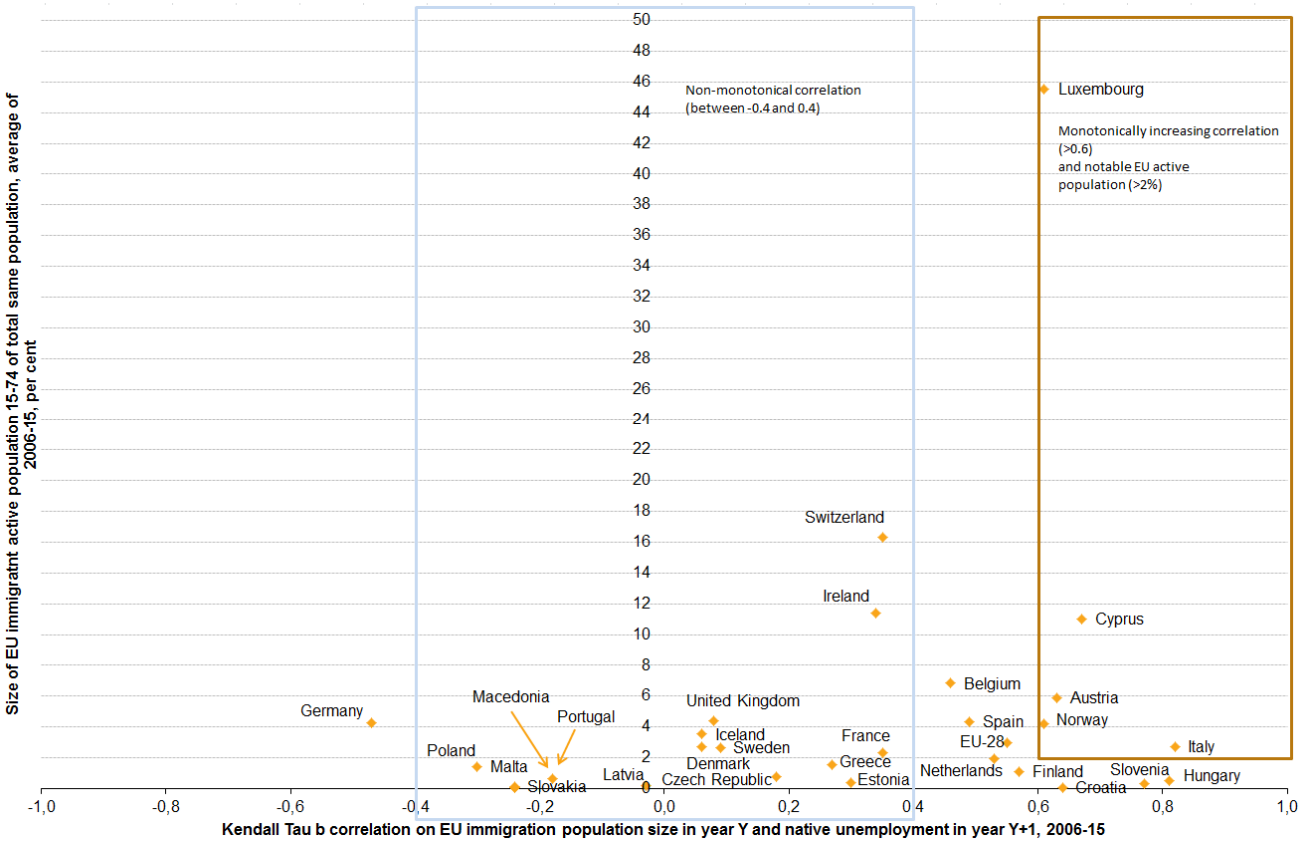


Population by sex, age, citizenship and labour status (1 000) [lfsa_pganws] in the Eurostat online data base http://ec.europa.eu/eurostat/data/database

With this background in mind we can go on to the first main topic, namely the correlations between immigration and unemployment. We use the Kendall Tau b correlation coefficient, for the correlation between the size of the immigrant active 15-74 population size in year Y and the native unemployment in year Y+1. The data set is the LFS for the time period 2006-2015. At country level, the result is the following.

Correlations on EU immigration and native unemployment, country findings

Chart 1: Correlation between EU immigrant population (active population 15-74) in year Y and native unemployment in year Y+1, Kendall Tau b, and size of the EU immigrant population, 2006-15, country level and EU-28 aggregate



Customized LFS data extraction

Even with the somewhat distorted scale due to the y axis value of Luxembourg, we see that most countries are placed low down on the y axis, and further that a substantial amount of them show no or only weak monotonic correlation (between -0.4 and 0.4 on the x axis).

As we see, Luxembourg, Cyprus, Austria, Norway, and Italy are the only countries which have the combination of an immigrant population size which could have some impact on the labour market (above 2 per cent of the active population), together with a non-negligible Tau

b value (above 0.6), meaning that the correlation for these countries is mostly monotonic. No countries have a high negative Tau b value. We will therefore take a closer look at these countries, and further contrast them to some selected countries with other properties (chart 2). Please note the different axis scales on the following graphs. This is not ideal for comparing results, but as the axis values vary a lot between the countries this is done to ensure readability.

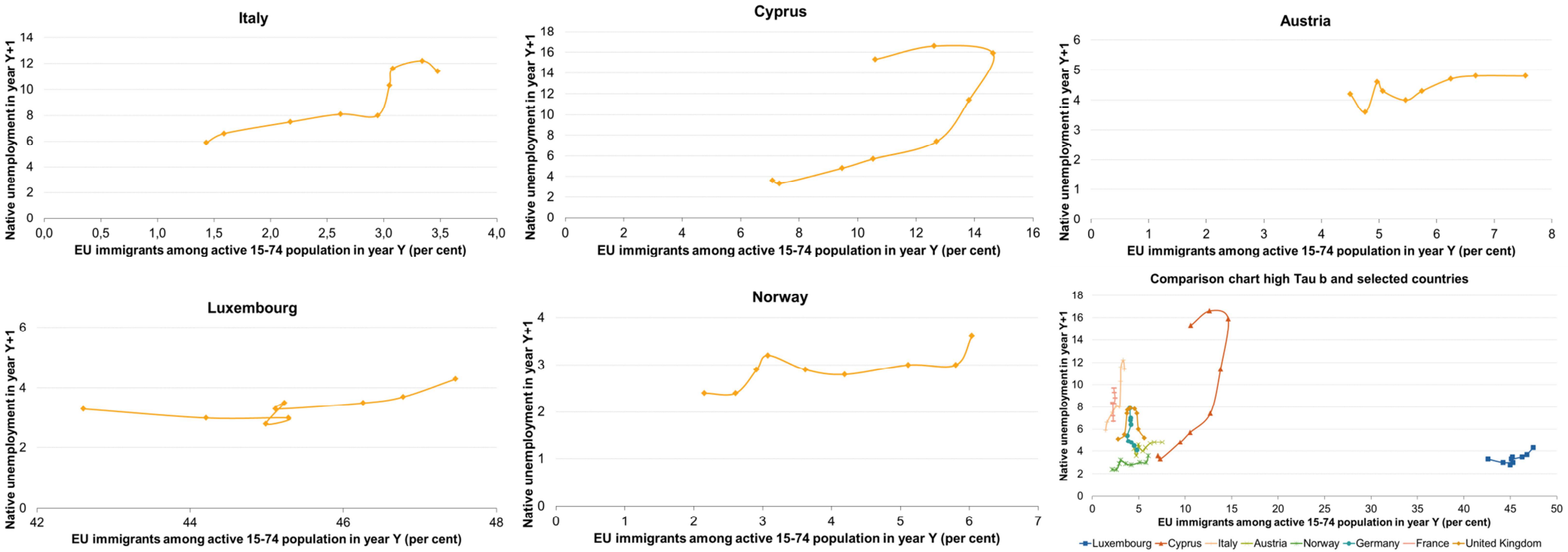
The highest Tau b value is in Italy, and chart 2 shows why this is the case: the native unemployment mirrors the movement of the immigration population for all but two data points. The second highest Tau b value in this group is found in Cyprus. The correlation is quite clear for a large majority the data points: when immigration increases, the native unemployment increases the year after, and when the immigration decreases, the native unemployment rate decreases the year after. There could be (and most probably are) many other variables involved in this trend, but these factors are outside the scope of this analysis. As far as this paper is concerned, the hypothesis of significant correlation between immigration and native unemployment is confirmed for this case.

The third highest Tau b value in this group is found for Austria. Here the development over time is notably more up and down at the start of the time period, and flattening out at the end, and operating inside a much less volatile labour market (only jumping between 3.6 and 4.8 per cent native unemployment). The Tau b value is nonetheless high enough that we can say that there is a monotonic correlation between the two.

On tied fourth place for the Tau b value we find Luxembourg and Norway, but with very different immigration population sizes. For Luxembourg we find, for the exception of the middle of the curve, a visible but not very pronounced correlated decline in immigration and in unemployment at the start of the time period, and the opposite at the other end of the curve. Norway shows a parallel rise in unemployment and immigrants, followed by a parallel decrease in unemployment and immigrants, with a flatter period in the middle, and then again a rise in both variables at the end of the time frame.

Providing comparison between the levels and developments in these countries, and a bit of context (lower right hand corner of chart 2), we see that for France the immigration population is stable (2.2 to 2.5 per cent), and a native unemployment rate which varies visibly (6.7 to 9.7). For the UK we see an almost symmetrical inverted U shape, meaning that native unemployment increased and then dropped, in a time period when the immigrant population increased continually (practically the definition of a non-monotonic correlation). For Germany we see a quite strong negative correlation, meaning that the immigrant population size increases continually while the native unemployment drops.

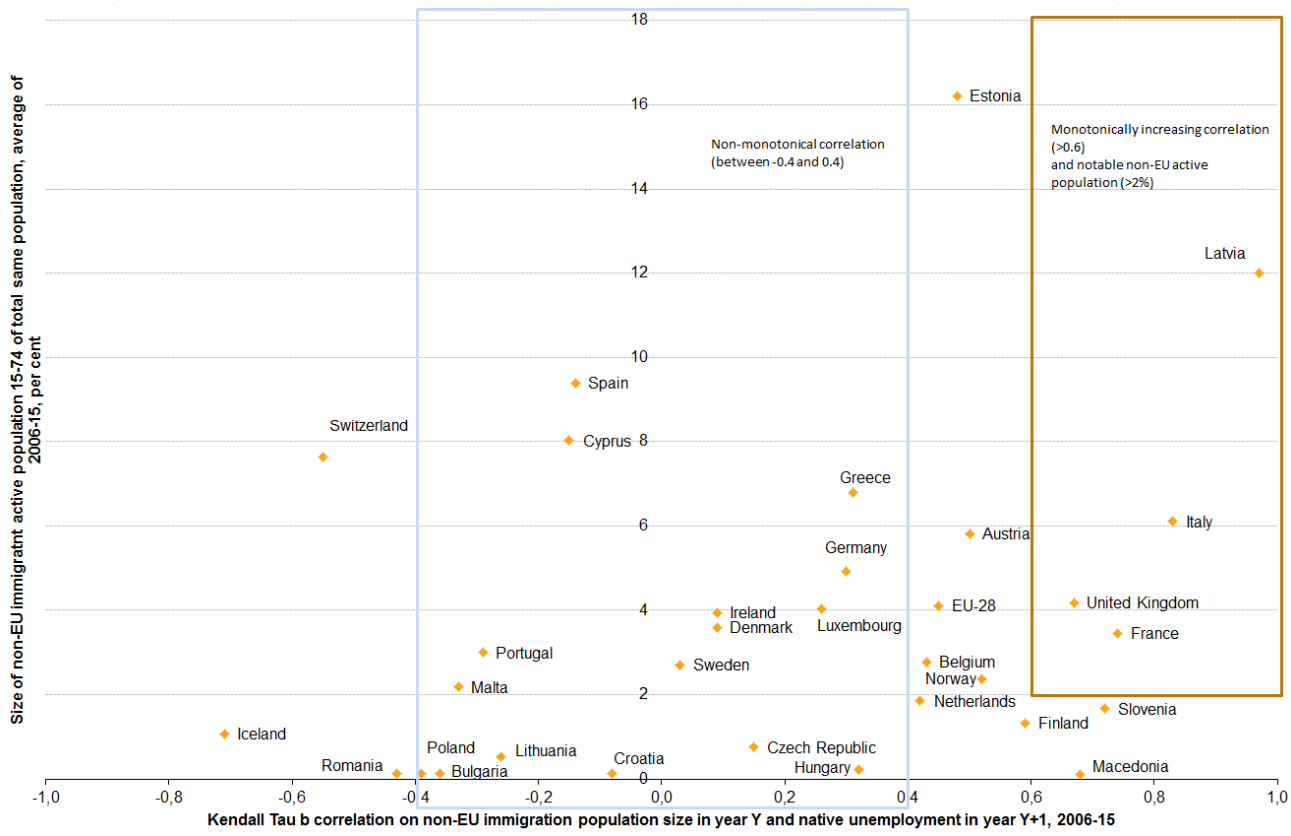
Chart 2: EU immigrant population (active population 15-74) in year Y and native unemployment in year Y+1, 2006-15, single years, high Tau b results and selected countries



Population by sex, age, citizenship and labour status (1 000) [lfsa_pganws] and Unemployment rates by sex, age and nationality (%) [lfsa_urgan], Eurostat online database

Correlations on non-EU immigration and native unemployment, country findings

Chart 3 Correlation between non-EU immigrant population (active population 15-74) in year Y and native unemployment in year Y+1, Kendall Tau b, and size of the immigrant population, 2006-15, country level and EU-28 aggregate

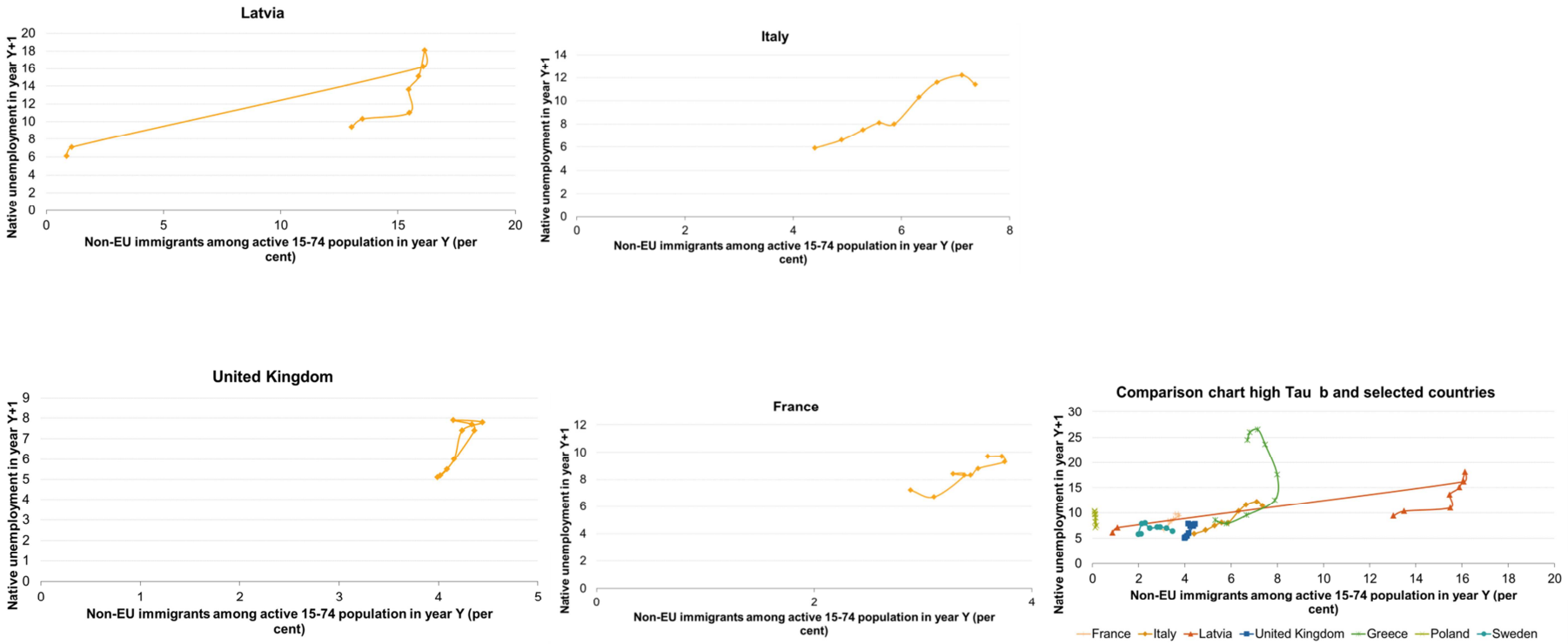


Customized LFS data extraction

Repeating the exercise, but now for non-EU immigrants, we see in Chart 3 that the countries of interest now are Latvia, Italy, the United Kingdom, and France. Please also note that y axis scale is different from the EU immigrant chart, as the non-EU immigration population does not have the massive outlier (Luxembourg) that was apparent in Chart 1.

The clearly highest Tau b value, at an almost impossible 0.97, is for Latvia. When looking closer on the country level graph in Chart 4, we find why: solving how to code recognized non-citizens means that the non-EU immigrant group increases from 12.7 thousand in 2007 to 183.8 thousand in 2008. Had it not been for this rather special case, the Tau value would have been lower.

Chart 4: Non-*EU immigrant population (active population 15-74) in year Y and native unemployment in year Y+1, 2006-15, single years, high Tau b results and selected countries*



Population by sex, age, citizenship and labour status (1 000) [lfsa_pganws] and Unemployment rates by sex, age and nationality (%) [lfsa_urgan], Eurostat online database

Also visible in Chart 4, the case of Italy is not due to recoding of data, and is still at a very high 0.83. This could of course just as well be explained by high non-EU immigration caused by war in the Middle East, and unemployment created by the finance crisis, occurring at the same time without having much, or even anything, to do with each other. However, the hypothesis for this paper is confirmed for Italy: native unemployment and non-EU immigration clearly correlate strongly.

For the UK we find a substantially more complicated curve. The curve starts and ends in the same situation (4 per cent immigrants, native unemployment a hair over 5 per cent), with both unemployment and immigrant population size first rising in parallel, then dropping in parallel, over these ten years.

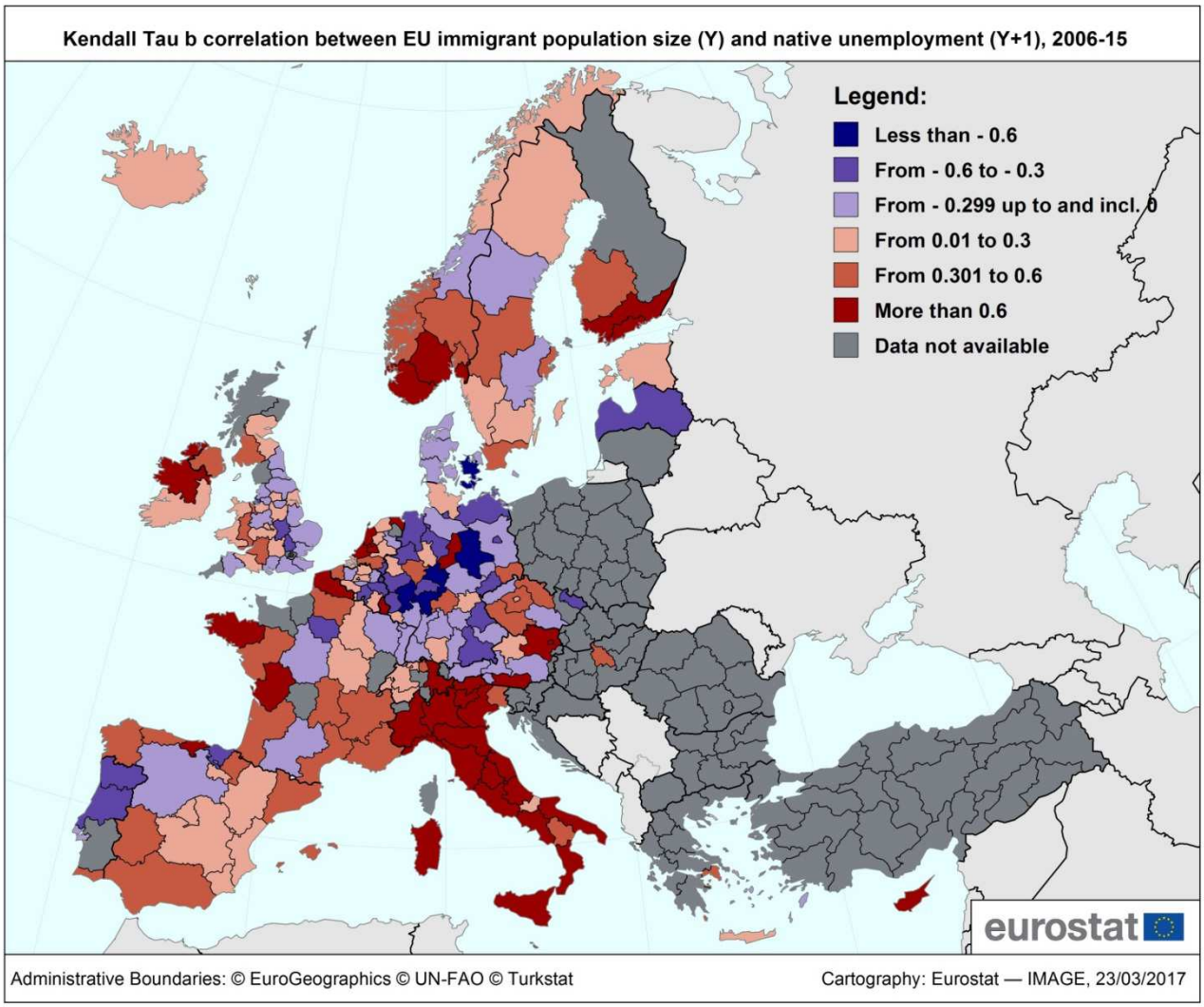
France shows a bit of back and forth as well, but the overall trend is clearly an increase both in unemployment and in the size of the immigrant population.

We also have a comparison for the non-EU immigration for the high Tau b countries and some other selected countries. Seen in context the curves of France and the UK are practically reduced to data points on top of each other, whereas Greece shows the largest movement that is not due to technical re-coding in the LFS. Italy's development over time is also clearly visible on this scale.

The added countries in this chart, Sweden and Poland, show here opposite developments from each other, with Poland having an increase in unemployment and a low, stable immigrant population, whereas Sweden has rather stable unemployment, but increasing immigration.

Correlations on EU immigration and native unemployment, region (NUTS2) findings

Map 3: Correlation between EU immigration (active population 15-74) in year Y and native unemployment in year Y+1, Kendall Tau b, 2006-15, NUTS2 level.



Customized LFS data extraction and analysis

For the display of correlations at regional level of analysis, we only keep the Tau b value for simplicity. That means that the immigration population size is no longer included as a separate variable, in order to keep the number of analytical groups (colours in the maps) to a manageable level. It is however implicitly included, as values under the publication threshold are represented by grey in the map.

Blue means negative correlation (increased immigration correlates with decreased unemployment), red means positive correlation (increased immigration correlates with increased unemployment). Stronger colour means stronger effect.

The most obvious trait of Map 3 is the strong east/west split on publishable data. The EU-immigrant population in Bulgaria, Hungary, Poland, Romania, Slovenia, Slovakia, Macedonia, and Turkey are below the threshold for what we reliably can say anything about on NUTS2 level. Greyed out regions are also, in some cases, results of changing NUTS2 borders over the time period we analyse (2006-15).

Overall, variation in correlation coefficients between the regions is largely contained inside countries, with Italy, Germany, Denmark, and the UK as prime examples of this.

Going more into detail, we see that France and the Iberian Peninsula, for the most part, show the same situation of medium correlations. Portugal's Centro and Norte regions, together with Spain's Castilla y Leon and Pais Vasco form a low negative correlation cluster, whereas the rest of these three countries are at the low positive side.

Another negative cluster is from Lorraine (FR41) eastwards through all of southern Germany into Steiermark (AT22). Vienna and its immediately surrounding area is on the high positive side, whereas we again find a continuous low or very low correlation zone from Severozapad (CZ04), through most of northern Germany, and into Denmark.

Netherlands and Belgium show a more mixed picture.

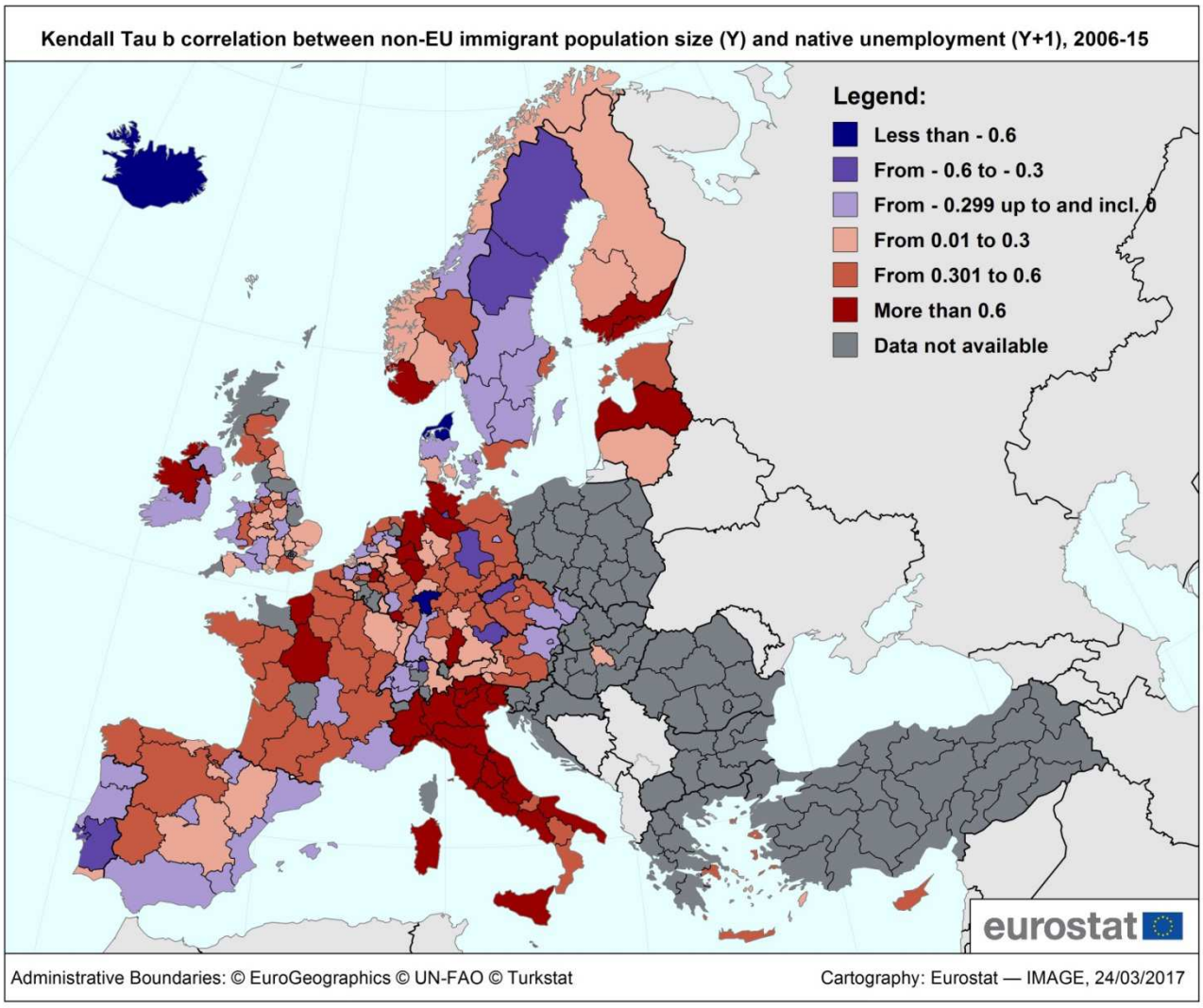
The United Kingdom varies between low negative on the east side, and low positive on the west side. Ireland has positive correlations throughout.

Italy clearly stands out as the country with the consistently high correlations for a large majority of its regions. The rest of the Mediterranean show mixed results, with Cyprus at high and Malta at low. Unfortunately, NUTS2 region changes make time series analysis for Greece difficult.

The Nordic countries are mostly on the low positive side, nonetheless with clusters of high positive around the capital regions of Finland and Norway.

Correlations on non-EU immigration and native unemployment, region (NUTS2) findings

Map 4: Correlation between non-EU immigration (active population 15-74) in year Y and native unemployment in year Y+1, Kendall Tau b, 2006-15, NUTS2 level.



Customized LFS data extraction and analysis

In Map 4 we find a bit different coverage of data (for instance for Finland and Lithuania), but one of the main pictures remains: immigration is much higher to the west than to the east.

Comparing the correlations of the non-EU immigrants to that of the EU immigrants, we find that France, Germany, and the United Kingdom have higher correlations (more red) for non-EU immigrants than for EU immigrants. We see further that correlations are overall lower (more blue) for Portugal, Spain, Cyprus, Ireland, Netherland, Sweden, Norway, and Cyprus. Italy stays the same.

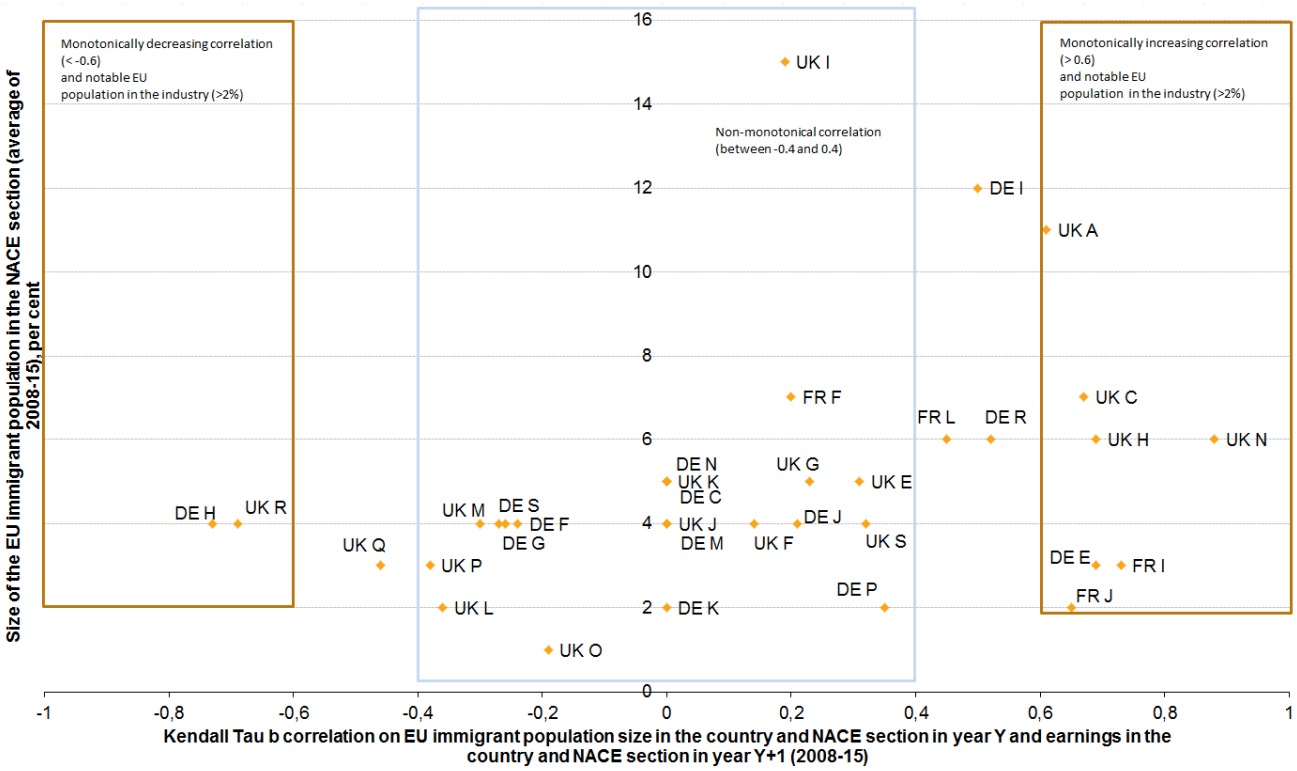
Correlations on EU immigration and earnings, by NACE sections

For the analysis of correlations between immigration and earnings we will use earnings statistics rather than the LFS INCDECIL variable, as deciles are not detailed enough to offer a useful analysis. The downside of this is that earnings data sets differ somewhat between countries, and that the country coverage is much smaller than for the LFS. This analysis requires annual earnings statistics for several years without breaks, distributed by NACE industry section, and enough immigrants to make a NACE industry section distribution of them meaningful. This chapter is therefore more on the experimental side, and we will only look at Germany, France, Portugal and the United Kingdom.

Immigration levels are measured in the same way as in the chapter on unemployment (coming from the LFS). Earnings data is taken from the published national results, and then indexised to make cross country comparison easier. 100 is the average (median for UK) of each year for the full survey. The data is distributed on NACE sections.

We see in Chart 5 that most of the correlations are low (the box in the middle of the chart). Only two industry sections, transportation and storage in Germany (DE H), and arts in the United Kingdom (UK R), show a clear negative correlation between immigration and earnings. These are therefore the only two cases in the data which have some basis for claiming that EU immigration precedes lowering of earnings. On the right hand side of Chart 5 we find seven opposite cases, that is, substantial correlations for that EU immigration is followed by increases in the industry wages. This is the case for the UK in agriculture (UK A), manufacturing (UK C), transportation and storage (UK H), administrative and support service activities (UK N), for Germany in water supply (DE E), and for France in accommodation and food service activities (FR I) and in information and communication (FR J).

Chart 5: Correlation between EU immigrant population (15-74) in the NACE section in year Y and earnings in the NACE section in year Y+1, Kendall Tau b, and size of the EU immigrant population, 2008-15, selected countries and industries.



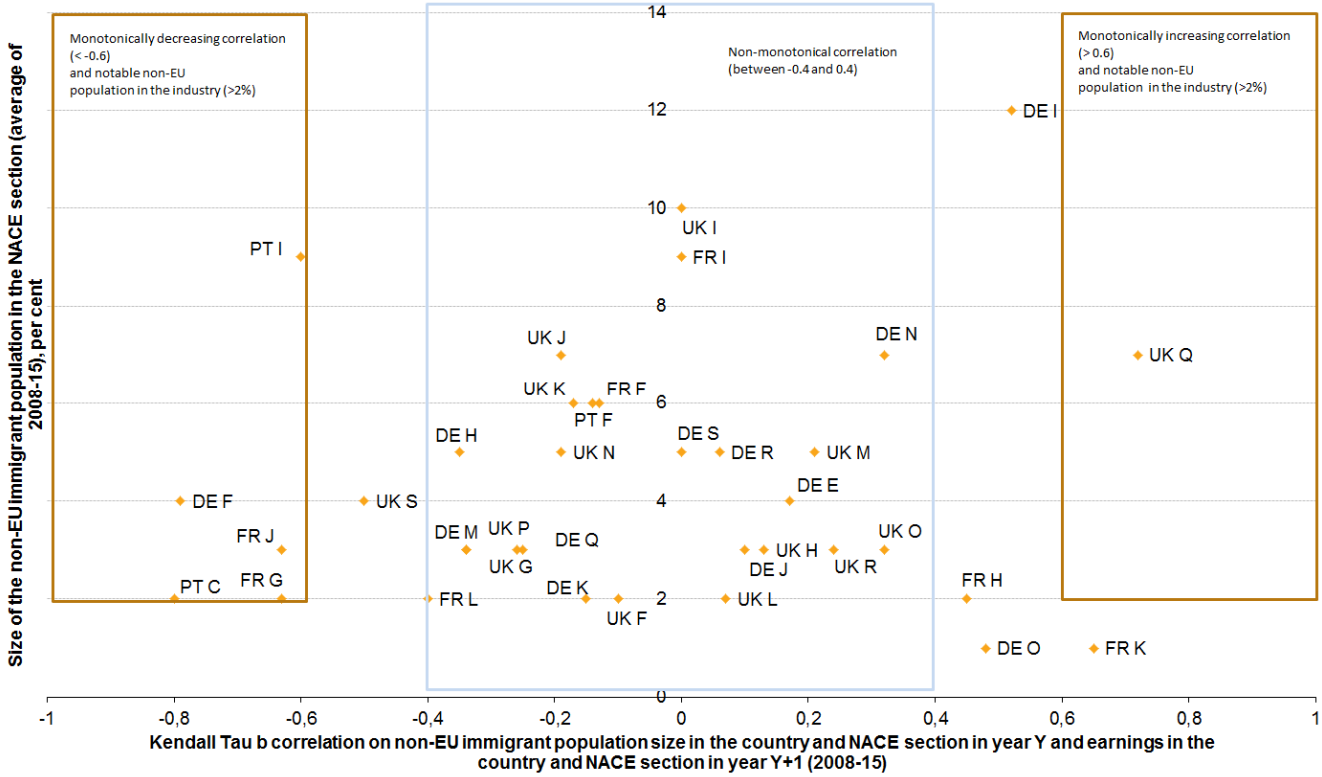
Customized LFS data extraction and analysis

DESTATIS, Durchschnittliche Bruttomonatsverdienste der vollzeitbeschäftigten Arbeitnehmer nach Wirtschaftszweigen und Jahren, https://www.destatis.de/DE/Publikationen/Thematisch/VerdienstArbeitskosten/Arbeitnehmerverdienste/ArbeitnehmerverdiensteLangeReihe.html

INSEE, Salaires mensuels moyens nets dans le secteur privé selon l'activité, http://www.insee.fr/fr/statistiques/2381338

ONS, Annual Survey of Hours and Earnings (ASHE), https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/earningsandworkinghours/datasets/ashe1997to2015selectedestimates

Chart 6: Correlation between non-EU immigrant population (15-74) in the NACE section in year Y and earnings in the NACE section in year Y+1, Kendall Tau b, and size of the non-EU immigrant population, 2008-15, selected countries and industries.



Customized LFS data extraction and analysis

DESTATIS, Durchschnittliche Bruttomonatsverdienste der vollzeitbeschäftigten Arbeitnehmer nach Wirtschaftszweigen und Jahren, https://www.destatis.de/DE/Publikationen/Thematisch/VerdienstArbeitskosten/Arbeitnehmerverdienste/ArbeitnehmerverdiensteLangeReihe.html

INSEE, Salaires mensuels moyens nets dans le secteur privé selon l'activité, http://www.insee.fr/fr/statistiques/2381338

INE, Average monthly earnings (€) by Geographic localization (NUTS - 2013), Economic activity (Section - CAE Rev. 3) and Employment size class; Annual , https://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_indicadores&indOcorrCod=0006917&contexto=bd&selTab=tab2

ONS, Annual Survey of Hours and Earnings (ASHE), https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/earningsandworkinghours/datasets/ashe1997to2015selectedestimates

For the non-EU immigrants we see that situation is different: most of the substantial correlations are on the left hand side of Chart 6, meaning that increased immigration is associated with decreased earnings. We find this for Portugal in accommodation and food service activities (PT I), and manufacturing (PT C), in construction in Germany (DE F), and

in France for information and communication (FR J) and wholesale and retail trade (FR G). The only case on the other side is human health and social work activities in the United Kingdom (UK Q). Nevertheless, the large majority of cases have low correlation coefficients.

Conclusions

Five out of the thirty three countries currently covered by the European Labour Force Survey show a notable positive correlation between EU immigration and native unemployment. For non-EU immigration we find this for four countries.

For most combinations of NACE industry sections by country there is a low correlation between immigration and earnings. There are eight cases of immigration being correlated with increased earnings, seven cases immigration being correlated with decreased earnings, and fifty five cases with low correlations.

None of the hypotheses under consideration in this paper have therefore been verified by the presented data.