LinkedinData

Giorgio Cristiano

3/25/2021

Impact of migration on GDP

As a next step, we would like to use the LinkedIN data to investigate how migration is impacting the growth of a country. To do so we are going to use the data provided by a collaboration between World bank and LinkedIN [citation needed]. To be consistent in our analysis we will carry out this analysis filtering only the countries whose WB Region is marked as "Europe and Central Asia".

While to get data on the countries GDP (normalized per capita), and GDP growth we are going to use the WDI API interface [citation needed]. To be consistent with the data provided by LinkedIN we are going to fetch the data between 2015 and 2019.

Now that we have the data that we need we can merge it in a single database that shows the migration flux among European countries, and GDPpc, GDP growth, and unemployment for both the source and destination countries. Each row contains information about a specific year, about the migration flux between a base and a target country. The migration flux is normalized according to the number of LinkedIN users in the base country. This number is positive when more people are arriving from, than leaving to the target country, and negative when the opposite happens.

Now that we have created the dataframe, we can start to investigate the relationship between the migration to a country and its wealth. The first step is to actually whether there is a trend in the destination of the migratory fluxes. To answer this question, Figure @ref(fig:migrationperyear) (left) plots a country's migration flux as a function of its GDP, grouped by year. From the plot we have removed Luxembourg, that is a clear outlier, possibly because of its very small population. What we note from this plot is that there seems to be a positive correlation between migration flux and destination country GDP, suggesting that, as expected, more people tend to emigrate to wealthier countries, where the quality of life expectancy is higher.

First of all we should check the significance of this result. We can do so through the p value. The highest value is 3×10^{-4}, which is well below the threshold of 0.05, implying that this correlation is significant. Then, as a next step, we could estimate the quality of the fit through the R^2 statistic individually for each year. This is in the range of 0.29 for the year 2016 to 0.36 for the year 2017. Meaning that whilst the fit is not perfect, it can explain more than the 25% of the data variance. Moreover, it can be noted how the coefficient of the fix seems to be decreasing with time. To have a better picture, Figure @ref(fig:migrationperyear) (right) shows the coefficient together with a 1

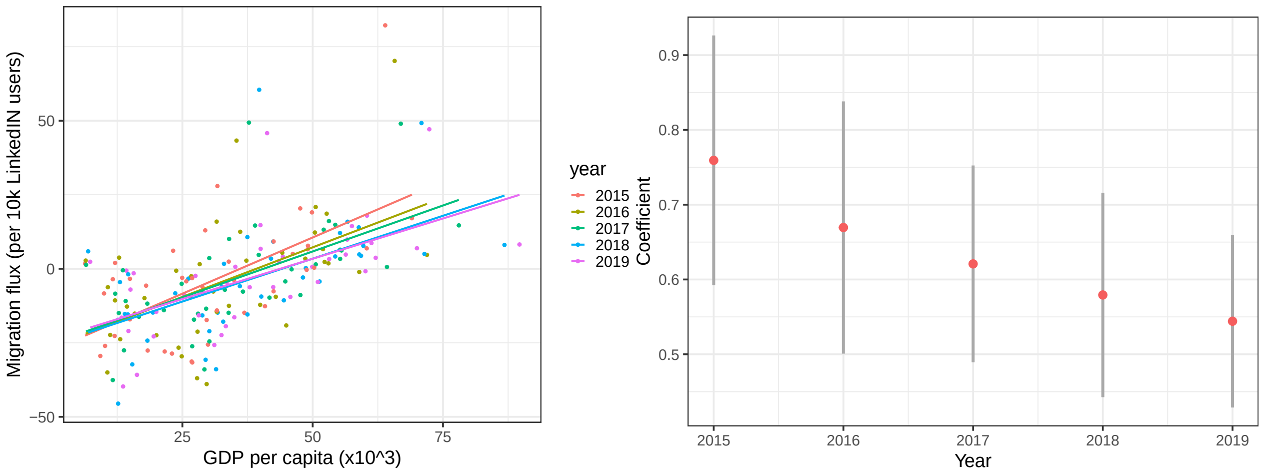


Figure 1: Relation between net flux migration of country X expressed as unit per 10k linkedIN users of that country and the country's GDP per capita. The data are grouped by year, and interpolated by means of linear regression.

std.err. per each year. From such figure we can evince that the decreasing trend seems to be within the 1 std.err. range, making it harder to distinguish from normal fluctuations. More data points will definitely be helpful in clearing up the picture.

To get more insights of the fit, we can plot the histogram of the residuals for the years with the best and worst fit, as shown in Figure @ref(fig:residuals_hist). As expected we can see that a better fit corresponds to less spread in the histogram of residuals.

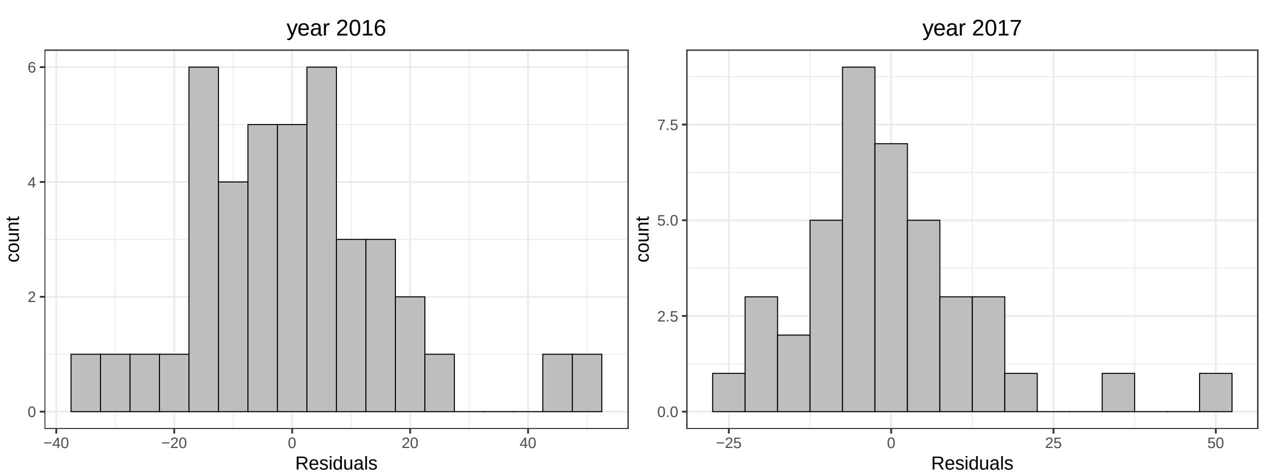
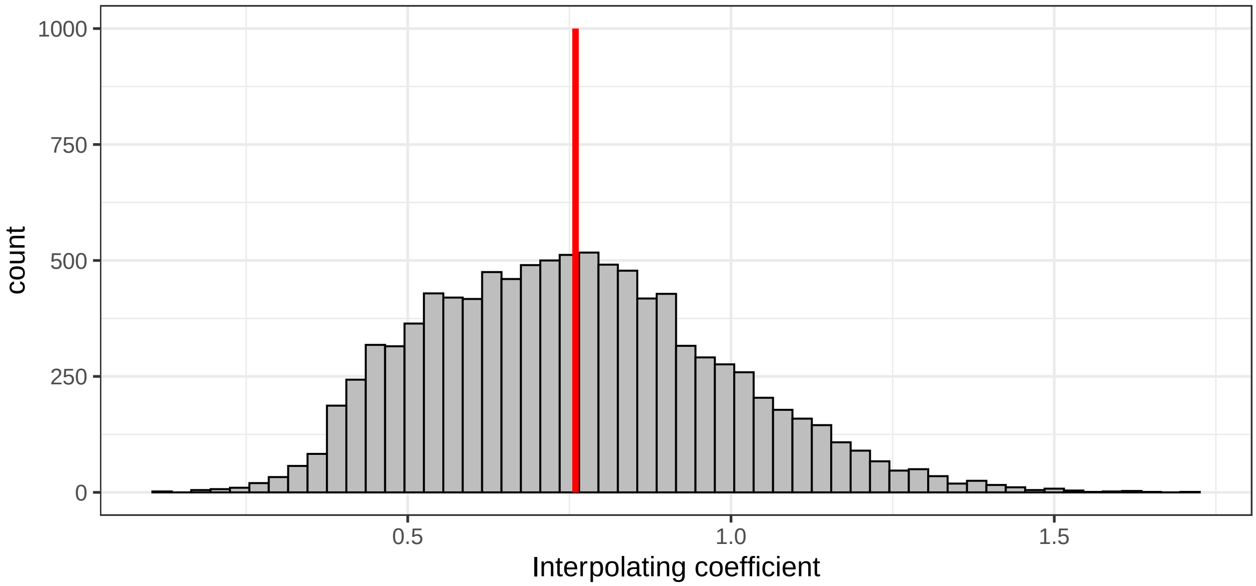


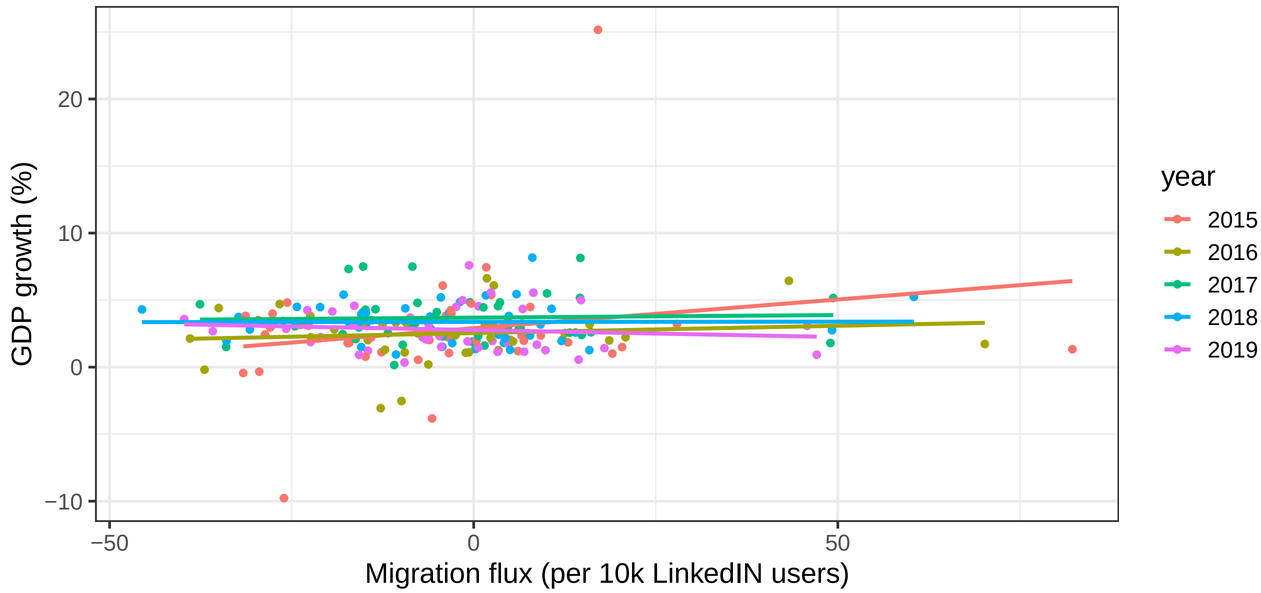
Figure 2: Histograms of residuals for the year with worst fit (left) and best fit (right).

Even though the dataset is quite small, since it includes one point per country per year, we can try to perform a bootstrap anaysis to check whether our data contains some bias. For simplicity, we will only perform this analysis on one single year; to be consistent, we will keep using the year 2015. Then, the bootstrap analysis will compute the linear regression coefficient over a limited dataset that allows repetitions. Here we have performed it for 10000 replicates, and the histogram showing the variation of interpolating coefficients is shown in Figure @ref(fig:bootModel).



From this histogram we can first of all note how the original coefficient is at the center of the gaussian, and how in 100% of cases, the interpolating coefficient is greater than 0. Thus we can comfortably state that there is a correlation between the migration flux and GDPpc of a country. At the same time, however, the distribution is quite wide spread, with a standard deviation of 0.23. Hence, whilst we can state the presence of a correlation, from this analysis alone, it is hard to quantify the strength of such correlation, as the addition of more data can significantly alter the results.

We have shown how wealthier countries tend to benefit from a larger working migration flux, it is interesting to study whether these countries are benefiting from this larger migratory flux. To do so we can plot the GDP growth as a function of the migration flux of a country. Again, as a first step, we will perform this analysis grouping the countries by year, and, as in the previous analysis, we will exclude Luxembourg, as it is a clear outlier. The results are shown in Figure @ref(fig:growthVSmigration).



This plot shows how there does not seem to be any correlation between GDP growth and migration, in fact the interpolating curves are very flat, and have very large residuals. This is confirmed by the p values, as they are all higher than 0.22. Moreover, the R^2 statistics are very low, up to 0.04. From this it is possible to conclude that, with only increase of production in mind, it should not be a country's first priority to be more appealing to foreign workers.

Industries preferred my migratory workers

To have a deeper look at how migration imapcts the growth of a country, we can also investigate whether a higher migration influx correlates to a larger growth of specific industries. To do so we are going to use again the linkedIN data that depicts the net gain (or loss) of members from (or to) a foreign country, for a specific industry, normalized to the number of linkedIN users, in that industry, in that country. These industries are gouped in different ISIC sections [citation needed]. The aim of this section is to verify whether net migration flux correlates to a larger growth of one of these sections. We are going to carry on this analysis for only one year, that is 2015.

Figure @ref(fig:growthVSmigration2) (left) depicts the data points grouped per ISIC section index, fitted through linear regression. Then, Figure @ref(fig:growthVSmigration2) (right) shows the slope of such interpolations, together with 1 std. error.

We can notice how the B, C and D sectors are lower than the others. These are Quarrying, Manufacturing and Energy respectively. On the other side, there is a stronger correlation with the sections F, I, R, which are Construction, Accomodation and Arts & Entertainment respectively. It is also important to notice that the industry growth due to foreigners can go up to a non-negligible 3%. With this in mind, incentivating a working migratory flux can help a country in develop certain industry sectors, that would otherwise be less populated.

The p values and R^2 values are showed in table (add Table). All p values are below the 0.05 threshold.Moreover, the sectors A and F have a very large std.err., meaning that there is probably

Table 1: R squared and p values for the fit of industry growth due to migration vs migration flux.

r.squared	p.value	isic_sector
0.4988701	0.0000267	I
0.4512041	0.0003253	N
0.7480841	0.0000004	R
0.4944769	0.0050222	D
0.2934236	0.0062579	Q
0.5836782	0.0000056	C
0.4717962	0.0409627	A
0.1314624	0.0414069	B
0.3149638	0.0005543	K
0.6102328	0.0000015	G
0.5486876	0.0000099	H
0.2201617	0.0077484	F

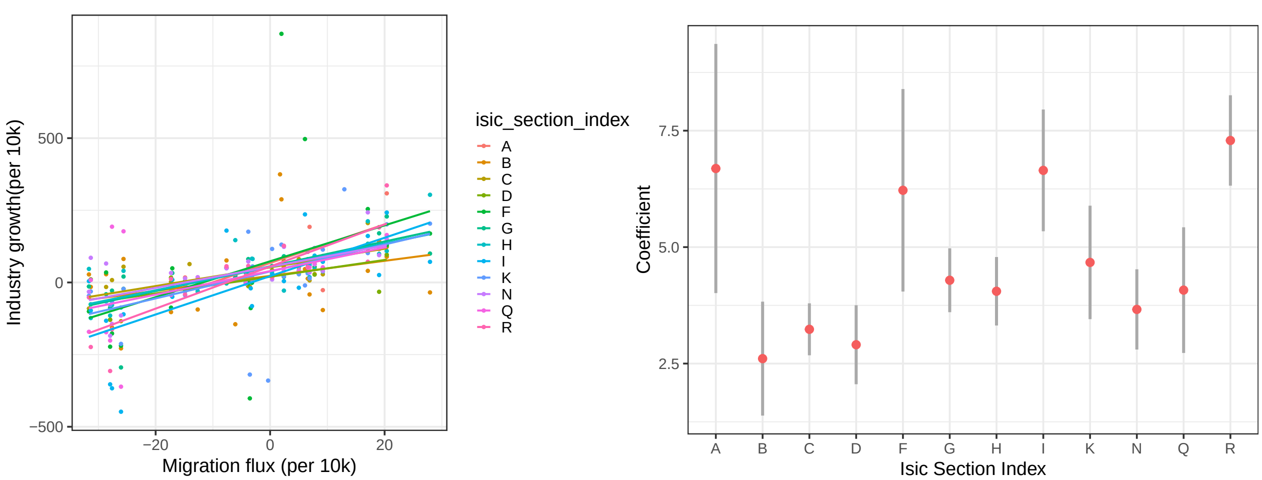


Figure 3: Industry growth due to migration as a function of net migration in the year 2019.

more variability across the countries, and a correlation with the migration flux alone is not enough.

Limitations

When doing such an analysis is it also important to identify where the limitations are. The first, most obvious one, is that this data has been normalized to the number of LinkedIN users for each country. We could not cross-validate this data with other employment datasets. Still, according to the source, this data has already been validated with 23 other external datasets. Moreover, these data do not include any differentiation such as gender or age. It would be interesting to investigate how migration fluxes differ for these categories.