Implications of employment indicators on GDP

Rosario Incandela, 20-901-252, rincandela@iis.ee.ethz.ch
Neerav Karani, 15-926-082, nkarani@vision.ee.ethz.ch
Giorgio Cristiano,17-800-491, giorgio.cristiano@iis.ee.ethz.ch

April 2021

1 Introduction

Unemployment is one of the toughest topics every government has to deal with. Especially after harsh economic crises (such as the one caused by the COVID-19 pandemic), unemployment tends to rise sharply: in the period between February 2020 - February 2021, 1 million people lost their jobs in Italy [6]. Moreover, certain sections of the population such as youth or females may be less protected by stable working contracts - thus, suffering higher unemployment rates. In the European Union, for instance, after the surge of COVID-19, the total unemployment rose by 1.465 million people from January 2020 - January 2021, corresponding to about a 1% increase while youths experienced a rising of 2% in unemployment in the same year [9]. Studies have shown that youth unemployment may have long-term and even irreversible consequences like permanent lowering of salaries or increased chance of subsequent unemployment in one's career [1], [3]. Such observations may motivate the need for introducing group-specific policies to improve the employment rate.

Another group-analysis that is relevant while considering employment indicators is the employment across different sectors. If it is observed that certain sectors contribute highly to the GDP, countries may be motivated to invest more in such sectors. For instance, in order to recover from the COVID-19 crisis, the European Union is going to release its largest stimulus package ever financed through the EU budget, amounting to a total of €1.8 trillion [8]. An sector-wise analysis of the employment influence on GDP is therefore highly pertinent, as it may influence policies concerning the distribution of such large economic resources.

Following this motivation, we analyze the implications of employment indicators on GDP. Specifically, we carry out the following analyses. (1) Firstly, we validate the hypothesis that total unemployment rate is negatively correlated with the GDP per capita. (2) Next, we carry out the correlation analysis for unemployment rates in different demographics. (3) Then, we analyze the effect of employment rates in different industry sectors on the GDP. (4) To understand better the two group-analyses mentioned above, we employ mediation to check if the effect of employment in a certain sector on the GDP is mediated by the unemployment in a specific demographic. (5) Finally, we extend study the effect on GDP of employment migration across-industry-sectors and across-countries.

2 Datasets, Tools and Methods

Data sets: We used data from two sources: (1) the World Bank database [2] and another database formed by a collaboration between the World Bank and the employment-oriented online service, LinkedIn [7].

Countries: Although both our data sources provide data from countries all over the world, we decided to restrict our analysis to European countries, so as to minimize the differences between the reliability of the measured data, and the methods of measurement. Accordingly, we considered the following countries for our analysis: Austria, Belgium, Bulgaria, Switzerland, Cyprus, Czech Republic, Germany, Denmark, Spain, Estonia, Finland, France, United Kingdom, Greece, Croatia, Hungary, Iceland, Italy, Lithuania, Luxembourg, Latvia, Netherlands, Norway, Poland, Portugal, Romania, Slovenia and Sweden.

Tools: All analysis was done in R and is available at `https://github.com/neerakara/SDS_project/tree/main`.

Methods: We used simple statistical tools: linear correlation (via the Pearson correlation coefficient), permutation tests (to evaluate the statistical significance of the results), bootstrapping (to compute uncertainty measurements), linear regression and mediation analysis (to probe the causal structure between the considered variables).

3 Analysis

3.1 Correlation between GDP per capita and total unemployment (2015 - 2019)

As the first step, we checked the correlation between the per-capita GDP and the total unemployment as a percentage of the total labour force. The data used for this analysis was obtained from the World Bank interface [2]. We considered the time period from 2015 to 2019, as both our data sources (Sec. 2) provide data in this duration.

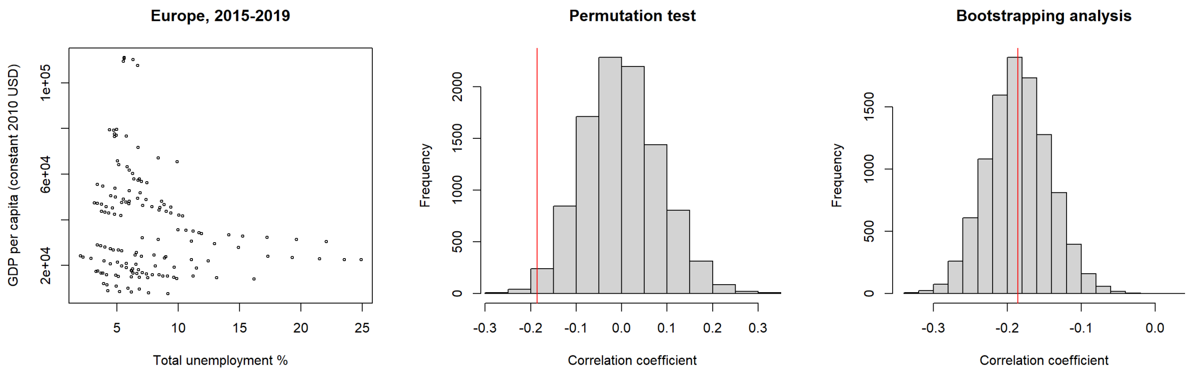

Figure 1: Correlation between the GDP per capita and the total unemployment rate (2015 - 2019). (a) scatter plot of data from all countries for all years in the considered duration, (b) histogram of correlations in different permutations of a permutation test, (c) histogram of the correlations in different bootstrapped datasets. In (b) and (c), the red line indicates the measured correlation value over the entire, unshuffled dataset.

Fig. 1a shows a scatter plot of the data from all countries (Sec. 2) for the mentioned duration. There is a small negative correlation between the total unemployment rate and the GDP per capita (Pearson coefficient is -0.186). A permutation test with 10^4 permutations (Fig. 1b) gave a p-value of 0.01, showing that the measured correlation is statistically significant. Further, a bootstrap analysis with 10^4 repetitions showed a mean correlation to be -0.185, with a standard deviation of 0.0425. Thus, although the measured coefficient is unbiased, there seems to relatively high uncertainty in the measurement. This indicates that the real correlation value for a larger data-set may be markedly different.

3.2 Year-wise analysis

The substantial uncertainty in the correlation measured in Sec. 3.1 could be due to multiple reasons. It may indicate that the relationship between the GDP per capita and the total unemployment is complicated, and perhaps dependent on several other variables. Also, the cumulative analysis over 5 years may be further complicating the relationship by introducing additional variations that may have occurred in different countries in different years. In order to alleviate this complication, we looked at the relationship between the GDP per capita and the total unemployment for each year separately. As there is no dependency of this year-wise analysis on the LinkedIn dataset, and to benefit from the additional data available via the World Bank interface, we included an extended time period: from 2005 to 2019.

A clear trend is visible in the results of the year-wise analysis (Fig. 2): the correlation between the total unemployment and the GDP per capita becomes consistently weaker from 2010 to 2019. However, the relationship is not entirely linear, as seen in the time period from 2005 - 2010. Furthermore, permutation testing shows

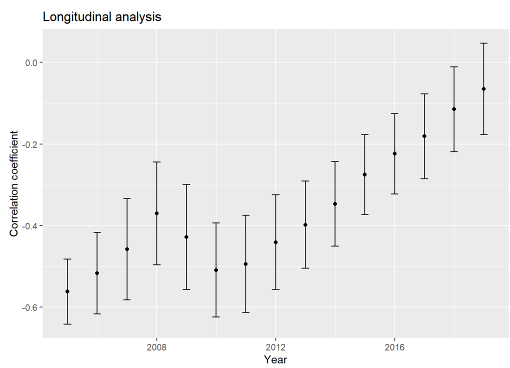

Figure 2: Year-wise correlation coefficients. Error bars show standard deviation computed using bootstrapping.

that the correlation values before 2016 are statistically significant, while the ones after 2015 are not. Based on this analysis, we concluded that there may be several time-dependent factors affecting the relationship between the GDP per capita and the total unemployment. A multivariate analysis including such additional factors is outside the scope of this project. Accordingly, we decided to focus on a single year for the further analysis. As the correlation is statistically significant before 2016, and we have LinkedIn data available from 2015 onwards, we carried out further analysis for 2015.

3.3 Correlation analysis: unemployment in different population groups (2015)

As a next step, we checked the correlation between the GDP per capita and unemployment in different groups in the population. Specifically, we considered the following 3 groups: unemployment in youth (between 15 and 24 years of age), unemployment in males, and unemployment in females. The results of this group analysis are shown in Table 1. It can be seen that the correlation values are quite similar for all the considered groups.

Group	Cor. Coef.	BT mean	BT std. dev.	p-value
Total	-0.275	-0.275	0.04	0.05
Youth	-0.287	-0.288	0.04	0.05
Male	-0.296	-0.296	0.05	0.05
Female	-0.251	-0.253	0.04	0.07

Table 1: Group analysis across population groups

3.4 Correlation analysis: employment in different employment sectors (2015)

Next, we considered the percentage of employment in three sectors: (1) agriculture (consisting of activities in agriculture, hunting, forestry and fishing.), (2) industry (consisting of mining and quarrying, manufacturing, construction, and public utilities (electricity, gas, and water) and (3) services (consisting of wholesale and retail trade and restaurants and hotels; transport, storage, and communications; financing, insurance, real estate, and business services; and community, social, and personal services). The goal of this analysis was to check if employment in a specific sector correlates with per capita GDP. The results of this analysis are shown in Table 2 and in Fig. 5 in the Appendix. The employment rates in agriculture and industry are negatively correlated with per-capita GDP, while employment in the services sector has a high positive correlation with the GDP per capita.

Employment sector	Cor. Coef.	BT mean	BT std. dev.	p-value
Agriculture	-0.507	-0.512	0.03	0.0001
Industry	-0.623	-0.622	0.04	0.0001
Services	0.730	0.731	0.03	0.0

Table 2: Group analysis across employment sectors

3.5 Mediation Analysis

In Sec. 3.4, we saw that the employment in some sectors negatively correlates with the GDP. To understand better the causal structure of this correlation, we carried out a mediation analysis to check if employment rates in specific sectors correlates with unemployment rates in specific groups, which in turn, negatively correlate with GDP. For instance, we saw that countries with a higher percentage of their population employed in 'industry' tend to have a lower per capita GDP. We checked if this correlation is mediated by increasing youth unemployment (perhaps because more manufacturing jobs are moved to lower wage countries outside Europe).

Three variables are involved in a mediation analysis [4]: an independent variable (I), a dependent variable (D) and a mediator (M) that can influence the relation between the first two variables. The analysis consists of 3 steps: (i) linearly regress D from I to obtain the *total effect*; (ii) linearly regress M from I; (iii) linearly regress D from I and M. The effect of I on D in the last step represents the *direct effect*, which, if mediation occurs should be drastically reduced or be insignificant. Using the *mediate* package, a summary of the indirect effect on D (namely *total effect* minus *direct effect*) is quantified as a single number by the Average Causal Mediation Effect (ACME). In Figure 3.5, a graphical model is shown: the red box is I (in our case, is the employment in a certain sector - agriculture, industry or services), the yellow box is D (the GDP), and the blue box is M (in our case, is the youth / female unemployment).

The results are shown in Table 3. A necessary condition for mediation to happen is that I has an effect on M. If this is not the case, then M is just a third variable that has an effect on D, but does not mediate the effect of I. This condition is not satisfied in our analysis, as the p-value of the regression between any of the employment sectors and the youth (or female) unemployment is much larger than 0.05. This means that our hypothesis is not verified and that we cannot conclude that unemployment of youths or females mediates the employment-to-GDP relation.

Sector (I)	I → D	p-value	Group (M)	I → M	p-value	I+M → D	p-value	ACME	p-value
Agriculture	-2330.97	0.00	Youth	0.64	0.14	-2128.0/-317.1	0.01/0.37	-202.98	0.32
			Female	0.25	0.25	-2181.9/-600.4	0.01/0.40	-149.06	0.61
Industry	-2238.58	0.00	Youth	-0.29	0.40	-2469.0/-796.4	2.47e-05/0.01	230.42	0.35
			Female	-0.19	0.26	-2557.5/-1675.2	1.70e-05/0.01	318.88	0.30
Services	1877.32	7.1e-06	Youth	-0.05	0.84	1850.5/-516.6	4.26e-06/0.05	26.83	0.75
			Female	0.02	0.87	1899.9/-1131.9	2.39e-06/0.03	-22.56	0.85

Table 3: Linear regression coefficients, p-values obtained from mediation analysis. For the column $I + M \to D$, coefficients for I and M from the joint regression are shown. ACME is the Average Causal Mediation Effect [4].

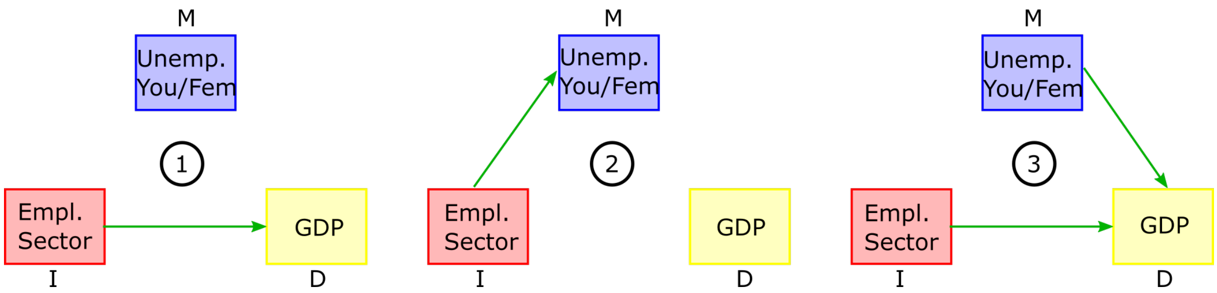

Figure 3: Graphical model of the mediation analysis. The green arrow represents the regression to be evaluated. Three steps: 1) I → D; 2) I → M; 3) I+M → D.

3.6 Migration Analysis using LinkedIn data

Next, we studied the relationship between GDP and employment migration, obtained from the world bank interface and LinkedIn respectively (see Sec. 2). To be as consistent as possible with the previous analysis, we selected countries marked as "Europe and Central Asia" in the LinkedIn data. By merging the two datasets, we had information about the migration flux, MF, (normalized by the number of LinkedIn users in the base country) among European countries, and the per-capita GDP, GDP growth and unemployment for source and destination countries. MF is positive when more people arrive to, than leave a particular country, and negative when the opposite happens.

3.6.1 Correlation between GDP per capita and migration flux (2015 - 2019)

A scatter plot of the MF and per-capita GDP is shown in Fig. 4a. A positive correlation is seen, demonstrating that more people tend to migrate to wealthier countries. A permutation test gave a p-value of 0.00046, implying that this correlation is statistically significant. A bootstrap analysis, performed for 2015, with 10^4 repetitions, showed that the measured correlation is unbiased (Fig. 4c). However, the std. dev. of 0.17, when compared to a mean coefficient of 0.54, indicates substantial uncertainty. The true correlation value may differ if we have more data points.

Next, we estimated the quality of the fit through the R^2 statistic individually for each year. This is in the range of 0.27 for 2016 to 0.35 for 2019, thus it can explain more than the 25% of the data variance. Further, the coefficient of the fit decreases with time (Fig. 4b). The decreasing trend seems to be within the 1-sigma range, making it harder to distinguish from normal fluctuations. Additional data may help in establishing a clearer year-wise relationship. Histogram of the residuals for the years with the best and worst fit, based on the R^2 statistic, are shown in Appendix Figure 7a,b. As expected, we see that a better fit corresponds to less spread in the histogram of residuals.

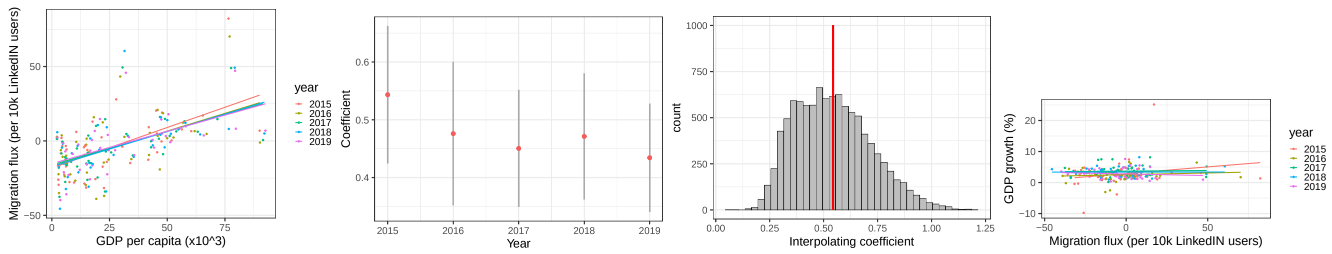

Figure 4: Relation between net flux migration per 10k LinkedIn users and the different GDP indicators. (a) Scatter plot of per capita GDP v/s migration flux. Data are grouped by year and interpolated via linear regression. (b) Variation of the fit coefficient over time (2015 to 2019). (c) Histogram of coefficients when performing a bootstrap analysis of migration flux vs GDP per capita of year 2015, with R = 10000. In red is depicted the interpolating coefficient of the original dataset. (d) Scatter plot of GDP growth v/s migrating flux for 2015 to 2019.

3.6.2 Correlation between GDP growth and migration flux (2015 - 2019)

Next, we carried out a similar analysis as in Sec. 3.6.1, using the GDP growth instead of the per-capita GDP as the dependent variable. A scatter plot (Fig. 4d) shows not much correlation - the interpolating curves are flat and have large residuals. This is confirmed by the p-values (all are higher than 0.22). Moreover, the R^2 statistics are very low (up to 0.04). From this and the previous subsection, one may conclude that workers tend to migrate to developed (high GDP) rather than developing (high GDP growth) countries.

3.6.3 Industry expansion deriving from migration flux

Finally, using industry-specific MF information from the LinkedIn data, we investigated whether a higher MF correlates with larger growth of specific industries. Industries are grouped in different ISIC sections [5]. We did this analysis for only one year, 2015. Fig. 5a depicts the data points grouped per ISIC section index, fitted through linear regression, while Fig. 5b shows the slope of such interpolations, with the error bars indicating the standard error. Growth in sectors like B (Quarrying), C (Manufacturing), D (Energy) shows low correlation with MF, while in some others such as F (Construction), I (Accommodation), R (Arts & Entertainment) shows high correlation. A higher MF can help certain sectors grow (in terms of their number of employees) by as much as 3%. Thus, incentivizing a working migratory flux can help countries develop certain industry sectors, that would otherwise be less populated. Some sectors (e.g. A, F) have a very large standard error, indicating high variability in the correlation in different countries. R^2 and p-values are shown in Table 4. All p-values are below the 0.05 threshold.

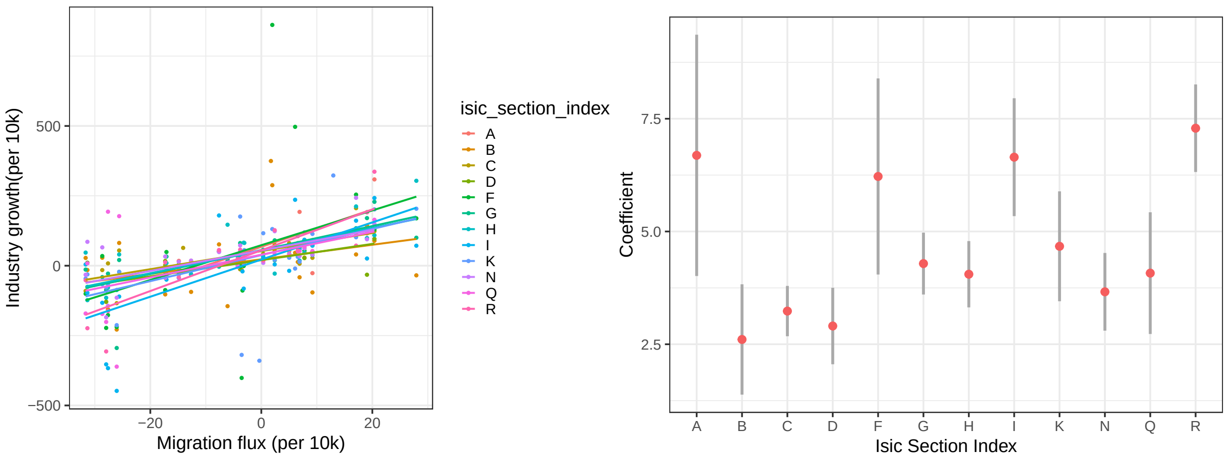

Figure 5: (a) Scatter plot of Industry growth v/s migration flux in 2015. (b) Correlation coefficients for different industry sectors.

isic_sector	A	B	C	D	F	G	H	I	K	N	Q	R
p-value x 10^6	41000	41000	5.5	5000	7700	1.5	9.9	26	550	320	620	0.42
R^2	0.47	0.13	0.58	0.49	0.22	0.61	0.54	0.49	0.31	0.45	0.29	0.74

Table 4: R^2 and p-values for the fit of industry growth due to migration vs migration flux.

4 Conclusions

In conclusion, the hypothesis that, if a higher percentage of a country's population contributes to its economic activity, the country's per-capita GDP would be higher was tested in 3.1. The analysis showed a negative correlation, thus validating our hypothesis. However, the bootstrapping analysis showed relatively high uncertainty. The group-analysis across population demographics (youth, male, female, in 3.3) showed a negative correlation too. Industry sector-wise analysis (3.4) showed that, for developed countries, more employment in the service sector is positively correlated with higher per-capita GDP, differently from the agriculture and industry sector. These correlations were not mediated by the unemployment in groups such as youths or females (3.5). Finally, we investigated the impact of migration on GDP. We saw that while wealthier countries tend to benefit from larger positive employment migration flux, such a flux does not correlate with GDP growth. Sector-wise migration analysis showed that migration may contribute to growth of specific industries such as Accommodation and Arts & Entertainment.

5 Discussion and limitations

We studied the effect of various employment-related indicators on the GDP. While allowing for some interesting observations, such an analysis is naturally limited in several ways. Firstly, the GDP is a macroeconomic variable that is affected by several inter-dependent factors - thus, the results of any bi-variate analysis should be treated with a grain of salt. This can be seen in the mediation analysis, where the mediator is actually another independent variable. Secondly, the average GDP per capita considered in this report may hide certain inequalities in the population. So, it may be prudent to consider other application-specific economic indicators. Thirdly, we limited the number of data points in the analysis, both in terms of geography as well as time. We looked only at European countries to level out economic or social differences, and considered data of only 1 year after noticing variation in the correlation over time. Further analysis can consider modeling behaviour over time, or including data from more countries. Fourthly, the LinkedIn data was normalized to the number of LinkedIn users for each country, so our analysis has an implicit assumption of similar reach and popularity of LinkedIn in all the considered countries. Finally, while our analysis could benefit by combining data from two data sources, it would have been interesting to study group-specific (age, gender etc.) migration tendencies had the LinkedIn data also included such information.

References

[1] W Arulampalam. "Is unemployment really scarring? Effects of unemployment experiences on wages". In: *The Economic Journal* (2001).

[2] World Bank. *World Bank database*. URL: `https://databank.worldbank.org/home.aspx`.

[3] Simon Burgess et al. "The class of 1981: the effects of early career unemployment on subsequent unemployment experiences". In: *Labour Economics* 10.3 (2003), pp. 291–309.

[4] Tingley D et al. *mediation:R Package for Causal Mediation Analysis.R package version 4.4.2*. 2013. URL: `http://http://CRAN.R-project.org/package=mediation`.

[5] *International Standard Industrial Classification of all economic activities (ISIC)*. `https://unstats.un.org/unsd/publication/seriesm/seriesm_4rev4e.pdf`.

[6] ISTAT. *Press Release 6th April 2021*. URL: `https://www.istat.it/en/archivio/256262`.

[7] Linkedin. *Linkedin and World Bank database*. URL: `https://linkedindata.worldbank.org/`.

[8] *Recovery plan for Europe*. `https://ec.europa.eu/info/strategy/recovery-plan-europe_en`.

[9] *Unemployment statistics*. `https://ec.europa.eu/eurostat/statistics-explained/index.php/Unemployment_statistics`.

Appendix

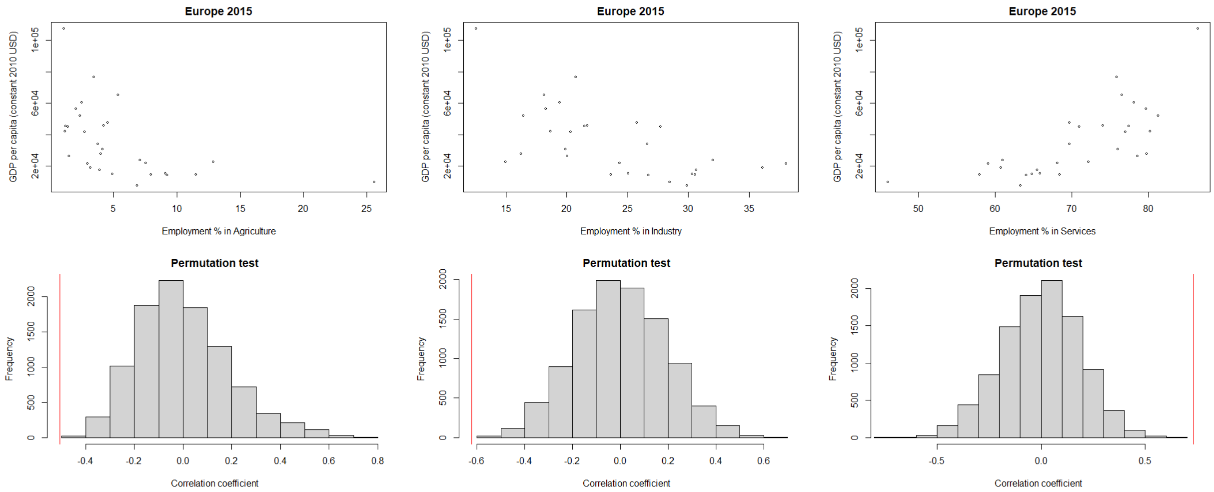

Figure 6: Correlation between the GDP per capita and the employment percentage in 2015 in different sectors: (a) Agriculture, (b) Industry and (c) Services. The top row shows scatter plots of the data, while the bottom row shows histograms of correlation coefficients obtained from permutation tests ($N = 10^4$).

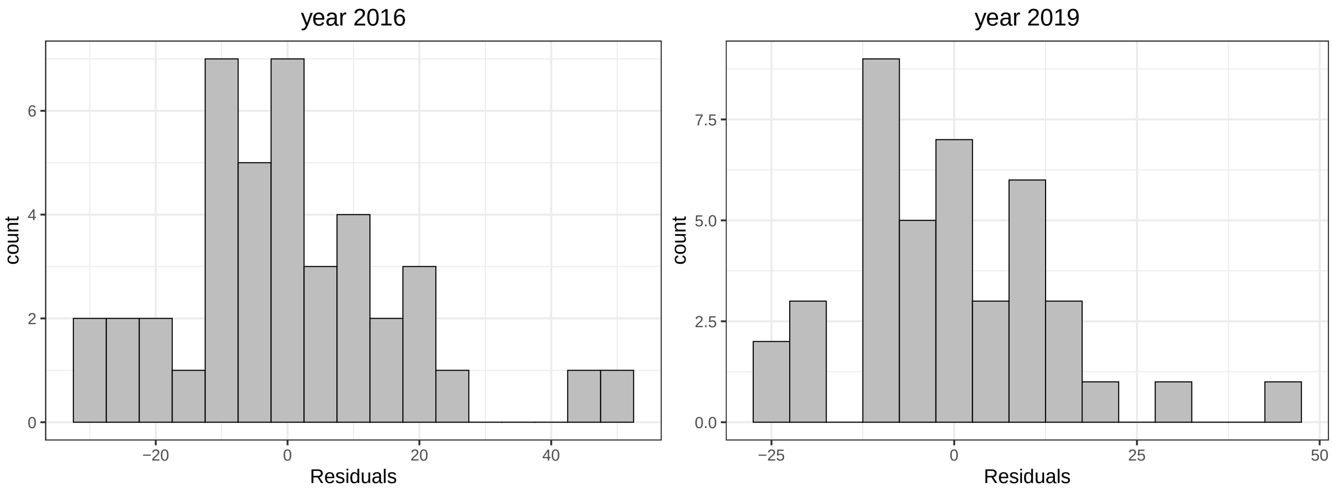

Figure 7: Histograms of residuals of correlation between GDP per capita, and migration flux for the year with (a) worst and (b) best fit, based on the R^2 statistic.