

**ESSnet Big Data II**

**Grant Agreement Number:**

<https://webgate.ec.europa.eu/fpfis/mwikis/essnetbigdata>

[<https://ec.europa.eu/eurostat/cros/content/essnetbigdata_en>](https://ec.europa.eu/eurostat/cros/content/essnetbigdata_en)

**Workpackage WPB**

**Implementation – Online Job Vacancies**

**Analyses of 1st CEDEFOP DATA (July-Oct. 2019)**

**Prepared by:**

**Swiss Federal Statistical Office**

**(Francis Saucy, Sophie Schmassmann, Alberto Columbano)**

**Date of present version: 24. July 2019**

Workpackage Leader:

Tomaž Špeh (SURS, SI)

e-mail address: tomaz.speh@gov.si

mobile phone: +38651672116

Contents

[Technical considerations 6](#_Toc15280436)

[Data description 6](#_Toc15280437)

[Number of records per country 6](#_Toc15280438)

[Duplicates 7](#_Toc15280439)

[Search for Patterns 8](#_Toc15280440)

[OJVs versus country population size 8](#_Toc15280441)

[Possible explanations for the lower rate of OJVs in CZ, ES, IT and PL 8](#_Toc15280442)

[Suggestions for testing those potential explanations 9](#_Toc15280443)

[GDP versus Population and OJVs (data July2018-March2019) 10](#_Toc15280444)

[Attempt of modelling OJVS versus population and GDP (NUTS1 model) 11](#_Toc15280445)

[Predictions for the 19 remaining countries for which no CEDEFOP data are available. 11](#_Toc15280446)

[Prediction maps 13](#_Toc15280447)

[Conclusions 13](#_Toc15280448)

[Additional remarks 13](#_Toc15280449)

[Further improvements (Sophie in work) 13](#_Toc15280450)

[Validation of predictions for Switzerland using external OJVs data 14](#_Toc15280451)

[X28AG OJVs versus NUTS1 prediction Model 14](#_Toc15280452)

[OJVS versus JVS 14](#_Toc15280453)

[Suggestions for further work and questions still to investigate 14](#_Toc15280454)

[Modelling OJVs at the regional level (NUTS2 model) 16](#_Toc15280455)

[Data exploration 16](#_Toc15280456)

[Modelling 17](#_Toc15280457)

[Conclusions : 21](#_Toc15280458)

[Further improvements (Sophie in work) 21](#_Toc15280459)

[Further improvements (Sophie in work) 21](#_Toc15280460)

[Overall summary of the study of the relationships between OJVs and socio-economical variables 21](#_Toc15280461)

[Comparisons between OJVs and Job vacancy Statistics (JVS) 22](#_Toc15280462)

[Rescaling the data 22](#_Toc15280463)

[Improving the comparability of the two data sources 23](#_Toc15280464)

[Subset of CEDEFOP data for NACE B to S 24](#_Toc15280465)

[Monthly vs quarterly data 24](#_Toc15280466)

[The case of Ireland 26](#_Toc15280467)

[Understanding the OJVs business models is a key issue 27](#_Toc15280468)

[Annexes: 28](#_Toc15280469)

[Detailed analysis of particular cases and extremely large OJVs 28](#_Toc15280470)

[Problems encountered in CEDEFOP Data 32](#_Toc15280471)

List of Figures

Figure 1 Distribution map of CEDEFOP OJVS per country 6

Figure 2 Relationship between OJVs and country population size 8

Figure 3 Same data as Figure 2 displaying distinct patterns for the two groups of countries 8

Figure 4 GDP versus population by country 10

Figure 5 OJVs versus GDP by country 10

Figure 6 Scatterplot of OJVs predicted (or fitted) by the model (vertical axis) against observed OJVs for the 13 countries of the CEDFOP data (log scales). 11

Figure 7 Prédictions of the model 12

Figure 8 Prediction maps based on the NUTS1 model (compare to Figure 1) 13

Figure 9 Comparison between X28 online vacancies collected for Switzerland with the Job vacancy statistics disseminated by the SFSO at the level of : 15

Figure 10 Plot of OJVs per country showing scatter range of observed values 16

Figure 11 Boxplots of OJVs per country. 16

Figure 12 Boxplots of OJVs (log scale) per country 17

Figure 13 Scatterplot of NUTS2 CEDEFOF OJVS (monthly average) against population size 19

Figure 14 Scatterplot of NUTS2 CEDEFOF OJVS (monthly average) against GDP 19

Figure 15 Standard plots of residuals 20

Figure 16 Scatterplot of fitted OJVs against observed values (log scale; corresponds to Figure 6). 20

Figure 17 Scatterplot of predicted OJVs against GDP (log scale; corresponds to Figure 7). 21

Figure 18 Graphical representation of data of Table 4 (log scales). 23

Figure 19 Graphical representation of the fidtance to a 1/1ratio of OJVS/ JVS with extreme values in opposite directions for IE and CZ. 23

Figure 20 Scatter plots of OJVs versus GDP for Italy 28

Figure 21 Scatter plots of OJVs (vertical axis) versus GDP for Germany 29

Figure 22 Scatter plots of OJVs (vertical axis) versus GDP for Spain 30

Figure 23 Scatter plots of average monthly OJVs(vertical axis) versus GDP for the United Kingdom. 31

List of Tables

[Table 1 Number of recorded OJVs per country (all months counfouded) 6](#_Toc15280495)

[Table 2 OJVs predictions for the 19 countries for which no CEDEFOP data are available (monthly average and 95%CI) 12](#_Toc15280496)

[Table 3 Comparison between X28 online vacancies collected for Switzerland with the Job vacancy statistics disseminated by the SFSO at the NACE2 level 15](#_Toc15280497)

[Table 4 Table of averaged CEDEFOP OJVs and Eurostat JVS figures. 22](#_Toc15280498)

[Table 6 Subset of NACE\_BS relatively to total OJVs per country. 24](#_Toc15280499)

[Table 5 Attempt to extract monthly OJVs coresponding to quarterly JVS for period 2018Q3 to 2019Q1. 25](#_Toc15280500)

[Table 7 Italy: NUTS2 OJVs and GDP ordered by descending OJVS showing the regional differences between northern and other NUTS2 regions 28](#_Toc15280501)

[Table 8: NUTS2 OJVs and GDP ordered by descending OJVS 29](#_Toc15280502)

[Table 9 Spain: NUTS2 OJVs and GDP ordered by descending OJVS 30](#_Toc15280503)

[Table 10 NUTS2 OJVs and GDP ordered by descending OJVS 31](#_Toc15280504)

# Technical considerations

The following analyses have been conducted during the period from July to October? 2019 using Jupyter notebooks on the BDTI platform. They refer to first set of CEDEFOP data presented in June 2019 at the physical meeting in Thessaloniki. Statistical modelling has been conducted using the open source software R.

# Data description

## Number of records per country

The dataset encompasses 64.3 millions of records. After exclusion of duplicate cases (query: SELECT COUNTRY, COUNT(DISTINCT GENERAL\_ID) as num\_job\_vacancy), the dataset displays 55.5 millions of unique records, or “Online job vacancy” (OJV) advertisements covering the period July 2018 - March 2019 and 13 countries (Table 1).

Table 1 Number of recorded OJVs per country (all months counfouded)

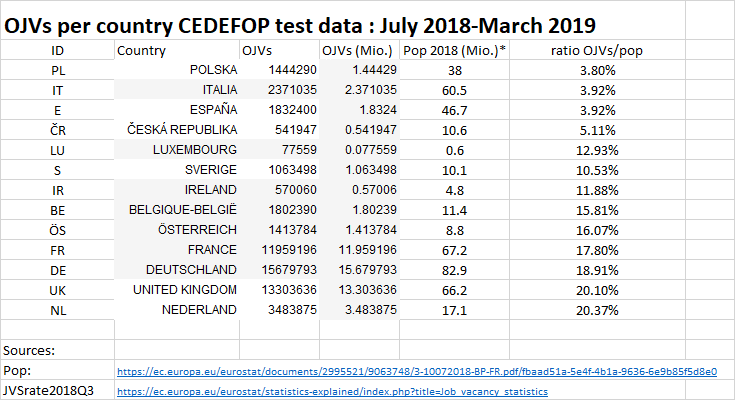
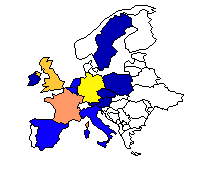
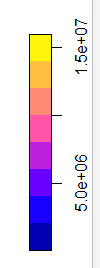
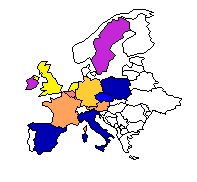
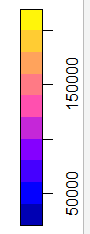


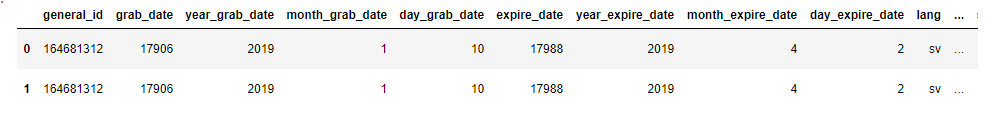
Figure 1 Distribution map of CEDEFOP OJVS per country

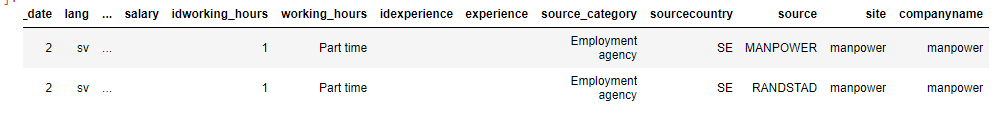
1. absolute figures b) relative to population size (OJVs per million)

## Duplicates

As indicated above, the dataset contains approximately 9 millions of duplicated records. Duplicates occur when advertisements are taken from another jobportal with identical values. In such cases two records are created with the same “general\_id” with different “idsource” and an indication of provenance in the field “companyname”. For instance, the case 164681312 (advertisement in Falun , SE) originally published by Manpower and republished by Randstad has been duplicated with indications of its first publication website.





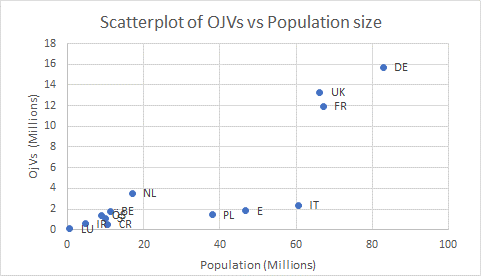
To be continued

# Search for Patterns

## OJVs versus country population size

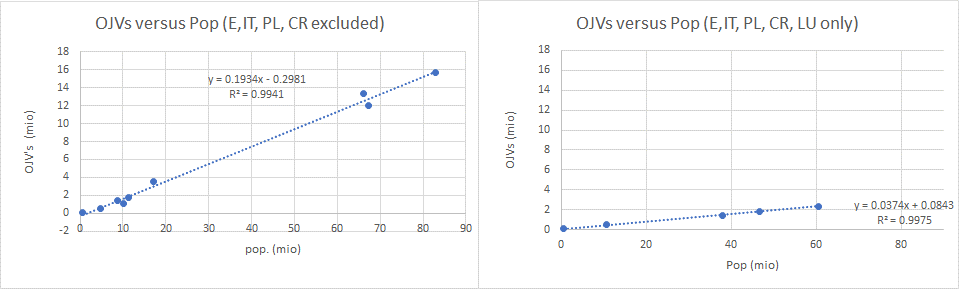
If OJVs have been collected using a similar approach in all countries and if Internet is used in a similar way among countries, one can expect a relationship between the total amount of OJVs and the size of the country. This is indeed the case as shown on Figure 2.

Figure 2 Relationship between OJVs and country population size



However, two distinct groups of countries seem to appear: one including most countries with a high ( >10%) rate of OJVs relatively to country population size and a second and smaller group including CR, ES, IT and PL with a lower rate of OJVs per million of inhabitants (<10%; Table 1). This pattern is clearly illustrated on Figure 3 with strong linear relationships (R2 >0.99 in both cases) which means that the variable “population size” in the regression model explains more than 99% of the variability in the OJS data among countries. It is unclear whether very small countries such LU belong to one or the other group.

Figure 3 Same data as Figure 2 displaying distinct patterns for the two groups of countries



## Possible explanations for the lower rate of OJVs in CZ, ES, IT and PL

There is a **bias or an artefact** in the collected data: only a part of the OJVs have been screened

1. The **OJVs channel** is not yet much developed in those countries
2. The difference reflects a **peculiarity of the labour market in** these countries
3. Others?

## Suggestions for testing those potential explanations

1. **Bias or artefact**
   * Plotting time series for all countries to detect anomalies;
   * compare the proportion of the scraped sources over time
   * search for duplicates
   * ask Tabulaex for peculiarities or scraping effort among countries
   * etc.
2. **OJVs channel**
   * Time series might display an increase in OJVs with time;
   * comparisons with other indicators of IT technologies and internet development among countries
   * etc.
3. **labour market**
   * Comparisons
     + with employment,
     + unemployment,
     + activity rates (proportion of persons in employment in the population)
     + Job vacancy rates
   * etc.

# GDP versus Population and OJVs (data July2018-March2019)

(R script: nuts1\_ojvs\_pop\_gdp.R)

An alternative explanation could be that OJVs mostly reflects some structural components of the economy of the various countries. This explanation is supported by the scatterplot of GDP versus population (Figure 4) in which the same four countries seem to exhibit a distinct pattern from the rest of the EU countries, as well as by the scatterplot of OJVs versus GDP (Figure 5) in which Spain and Italy also seem to display to a different pattern.

Figure 4 GDP versus population by country

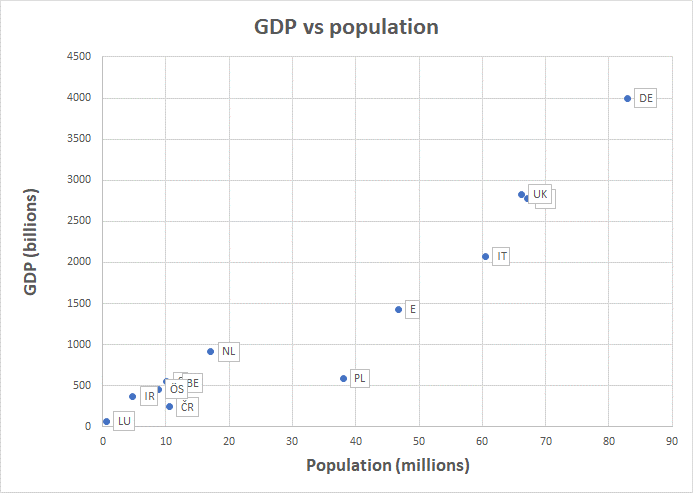
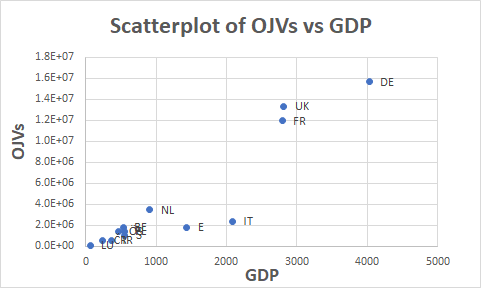


Figure 5 OJVs versus GDP by country



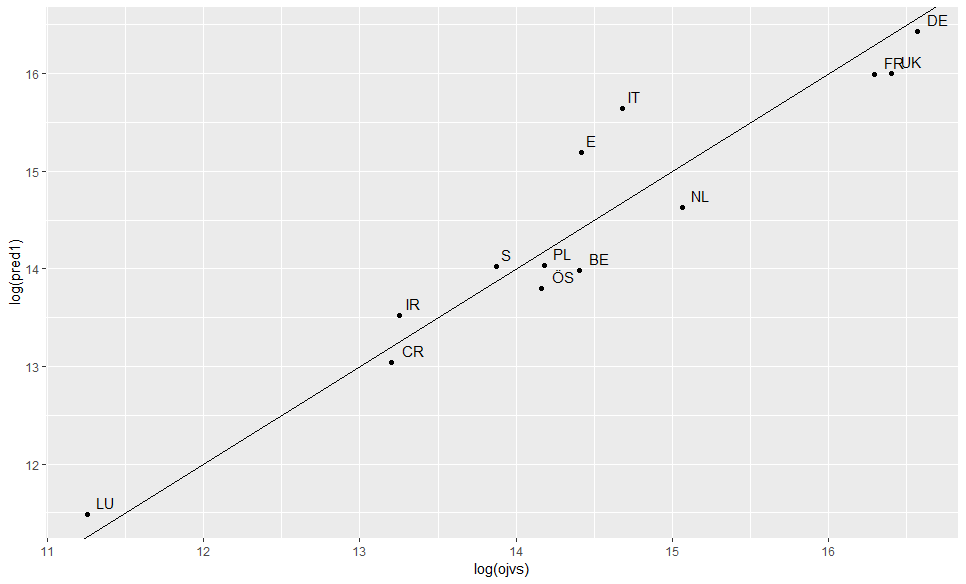
## Attempt of modelling OJVS versus population and GDP (NUTS1 model)

As suggested above, there seems to be a strong relationship between OJVs and population size and GDP at the country level. This relationship has been tested in a modelling approach by regression of OJVs against these two variables using the software R with the following results:

1. the regression of OJVs against Population size shows a significant relationship with an adjusted squared R of 0.66 indicating that population size explains 66% of the variance of OJVs (adj. R2=0.78 for log transformed data).
2. the regression of OJVs against GDP returns an even stronger relationship : adj. R2=0.85 (adj. R2 =0.90 for log transformed data).
3. the regression of OJVs against Population size and GDP returns an adj. R2 of 0.87 ((adj. R2=0.88 for log transformed data).

Figure 6 displays graphically the relationship between observed and fitted OJVs for the 13 countries for which CEDEFOP data are available. The graph shows that the data for all countries are aligned on a straight line (log scale) with dots for CR, PL, E and IT no more forming a particular cluster. Predictions for E and IT seem to be slightly overcorrected. Perhaps additional variables might help improving the fit for these countries.

Figure 6 Scatterplot of OJVs predicted (or fitted) by the model (vertical axis) against observed OJVs for the 13 countries of the CEDFOP data (log scales).



### Predictions for the 19 remaining countries for which no CEDEFOP data are available.

The results (Figure 7) of predicted OJVs (vertical axis) against GDP (horizontal axis) show a similar pattern of predictions for the 19 remaining EU countries (Figure 7b), as compared to the predictions for the 13 countries with CEDEFOP data (Figure 7a). All dots of the two groups of predicted OJVs against GDP are aligned on the same straight line. In order to assess the adequacy of the model, the predicted OJVs for the 19 countries of Figure 7b could be tested against actual data (like in Figure 6), once CEDEFOP data are available for those countries. These predictions with upper and lower 95% confidence intervals are given in Table 2.

Figure 7 Prédictions of the model

|  |  |
| --- | --- |
| 1. fit of the model for the 11 countries of the CEDEFOP data : predicted OJVs (vertical axis) against GDP (horizontal axis) | b) predictions of numbers of OJVS for the 13 remaining EU countries : predicted OJVs (vertical axis) against GDP (horizontal axis) |

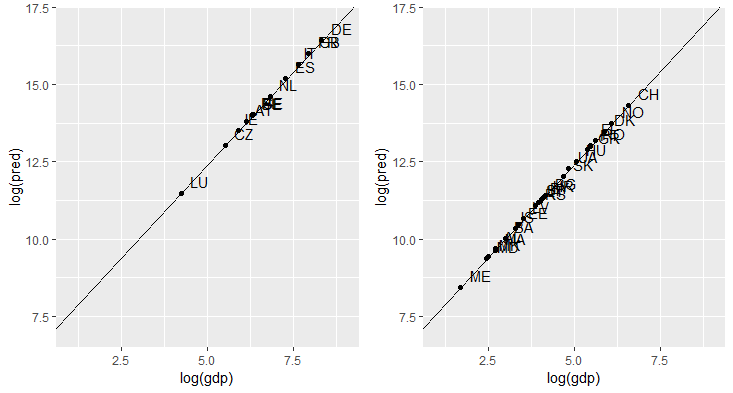
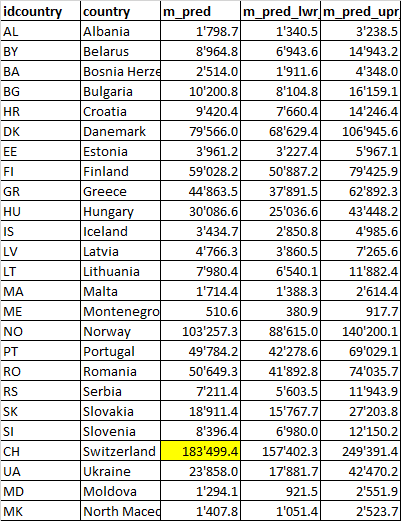


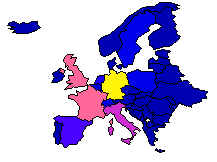
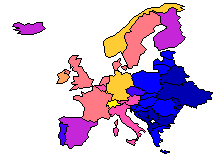
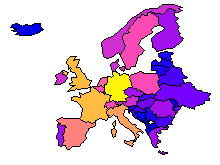
Table 2 OJVs predictions for the 19 countries for which no CEDEFOP data are available (monthly average and 95%CI)



### Prediction maps

Figure 8 Prediction maps based on the NUTS1 model (compare to Figure 1)

1. Predicted OJVs (abs. values) b) pred- OJV relative to pop. size c) pred. OJV relative to GDP

### Conclusions

The following conclusions can be drawn from the modelling exercise at the national level :

1. a strong relationship exists between OJVs, country size (population) and national economic activity expressed as GDP
2. the CEDEFOP data collected as the European scale seem to offer a promising basis for developing pertinent economic indicators
3. Model predictions have been drawn which could be compared to actual observations for the 19 remaining countries for which no CEDEFOP data are available
4. On log scale GDP and population size offer a reasonable predictive power for OJVs
5. On the same basis OJVs could offer a tool for developing flash estimates of GDP among EU countries

### Additional remarks

1. all calculations have been conducted using the summation of all OJVs per country collected over a nine-month period July2018-March 2019
2. they were regressed against annual population and GDP data.
3. Predictions should be multiplied by a factor 4/3 in order to get annual figures for OJVs or divided by 9 for average monthly figures

### Further improvements (Sophie in work)

1. Replace population size by employment (jobs, occupied posts or proxies such as employees (STS220), or persons in employment from LFS)
2. Produce a model with intercept = 0 (a country with pop and GDP = = 0 would have 0 OJVs)

## Validation of predictions for Switzerland using external OJVs data

Among those 19 countries, Switzerland is remarkable, because data for Switzerland are available from an external source, i.e. X28 AG, a Swiss private company which collects OJVS from websites of all individual companies on one hand and from largest job portals in Switzerland on the other hand. The Swiss case offers therefore a unique opportunity to :

1. compare the actual numbers of vacancies advertised by enterprises themselves against the official Job Vacancy Statistics (JVS) published by the SFSO on the basis of a classical survey,
2. to calibrate OJVs against official JVS

X28 AG OJVs versus NUTS1 prediction Model

According to the press releases of “Jobradar” based on X28 AG data for the period 2018Q3-2019Q1 corresponding to the CEDEFOP dataset (July 2018-March2019), an average of 187,400 OJVs have been collected in Switzerland, a figure very close to the prediction of 183,500 produced by the NUTS1 model (Table 2).

OJVS versus JVS

Among the overall 187,400 OJVs reported in Jobradar (<https://www.jobagent.ch/jobradar> ), 82,100 OJVs (44%) were directly collected from websites of individual enterprises and can therefore be compared tot he SFSO official Job Vacancy Statistics. After cleaning for differences in the NACE cover (only NACE sections B-S are covered by the JVS), we observe here again a very close concordance between the two sources, with 74,300 and 75,700 vacancies in the SFSO JVS and X28 AG data, respectively, the X28 AG estimates beeing included within the limits of th 95% confidence intervals of the SFSO statistics (±5,100). In addition, similar conclusion can be drawn for the estimates at the NACE2 and NUTS2 regions (Table 3; Figure 9).

Suggestions for further work and questions still to investigate

It would be valuable to study in more detail the X28 AG with respect to the predictions of the NUTS1 model in order to determine how they compare to the CEDEFOP source. In particular, it would be interesting to understand the relationships between OJVS from job portals, agencies of temporary work and websites of individual entreprises and how each category contribute to the total OJVs amount.

Table 3 Comparison between X28 online vacancies collected for Switzerland with the Job vacancy statistics disseminated by the SFSO at the NACE2 level

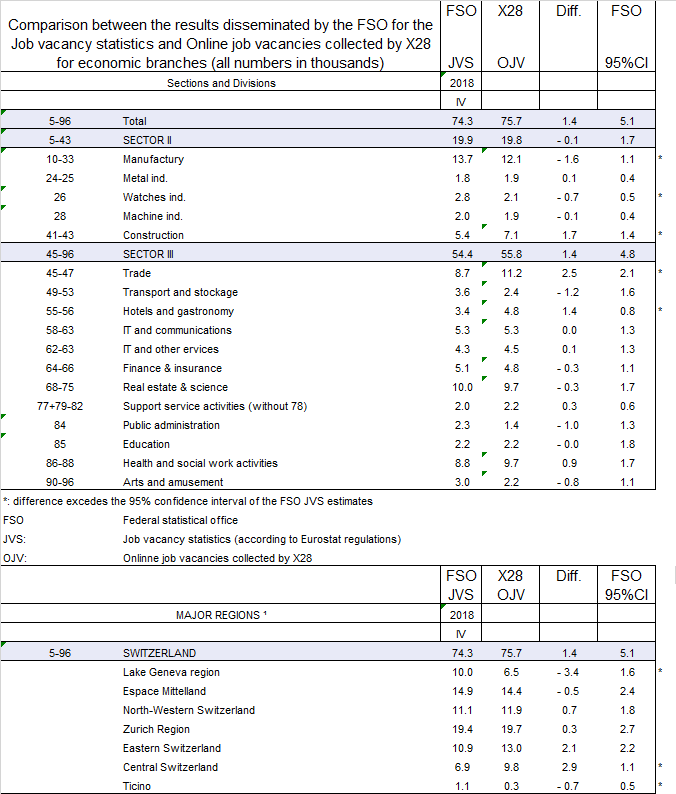
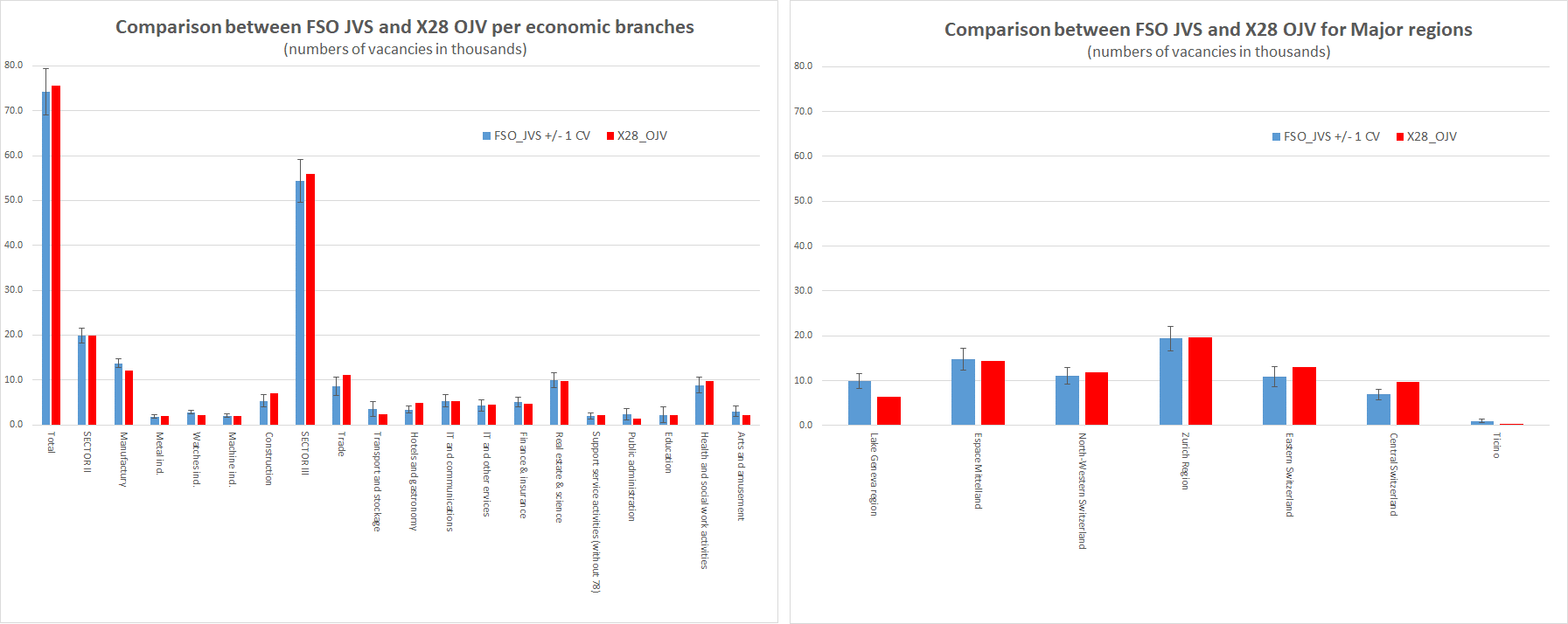


Figure 9 Comparison between X28 online vacancies collected for Switzerland with the Job vacancy statistics disseminated by the SFSO at the level of :

1. Economic branches b) Large regions



# Modelling OJVs at the regional level (NUTS2 model)

## Data exploration

Boxplots of regional OJVs per country are given in Figure 11After a log transformation of the data another set of outliers appears, cases with figures smaller than 1.5 times the interquartile range below the 25th percentile for three countries (ES, UK and FR). More detailed data for these countries are given in detail in the Annex section.

Figure 12 Boxplots of OJVs (log scale) per country

Legend : horizontal bars: median; coloured boxes ;25th and 75th percentile; vertical bars: range encompassing 1.5 times the interquartile range (below the 25th or above the 75thpercentile); red dots: mean; coloured dots: oberserved values.

. Outliers (or extreme values exceeding 1.5 times the interquartile range above the 75th percentile shown by the range bars) occur mostly in the largest EU countries (ES, IT, UK, DE and FR). Ireland (2 regions) and Luxembourg (1 region) are excluded. Three groups of countries appear on this graph where countries are ordered with increasing median values of OJVs :

1. the four countries mentioned earlier as countries with less than average OJVs numbers (CZ, ES, PL and IT)
2. a group of three countries with larger than average OJVs numbers (NL, DE and FR)
3. a group of four countries with intermediate OJVs numbers (SE,AT,UK and BE).

Figure 10 Plot of OJVs per country showing scatter range of observed values

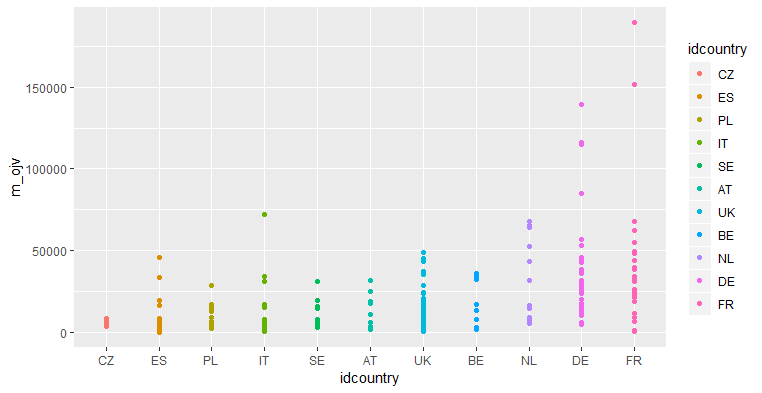
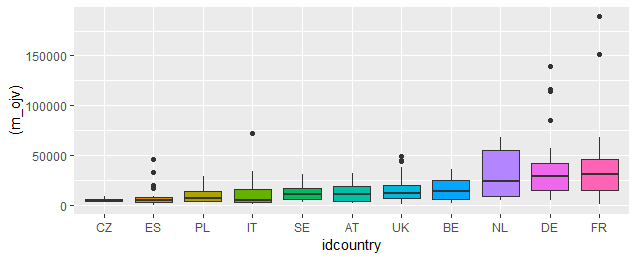


Figure 11 Boxplots of OJVs per country.

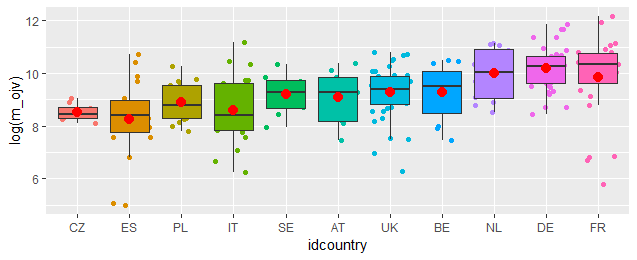
Legend : horizontal bars: median; coloured boxes ;25th and 75th percentile; vertical bars: range encompassing 1.5 times the interquartile range (below the 25th or above the 75thpercentile); black dots: outliers



After a log transformation of the data another set of outliers appears, cases with figures smaller than 1.5 times the interquartile range below the 25th percentile for three countries (ES, UK and FR). More detailed data for these countries are given in detail in the Annex section.

Figure 12 Boxplots of OJVs (log scale) per country

Legend : horizontal bars: median; coloured boxes ;25th and 75th percentile; vertical bars: range encompassing 1.5 times the interquartile range (below the 25th or above the 75thpercentile); red dots: mean; coloured dots: oberserved values.



## Modelling

Similar calculations have been conducted at the NUTS2 regional level using

(Rscript: nuts2\_ojvs\_pop\_gdp.R)

1. the CEDEFOP OJVs data for the 13 countries against auxiliary variables (regressors) drawn f
2. auxiliary variables (regressors) were drawn from the following Eurostat website:

[https://ec.europa.eu/dashboard/publisher/vislet.html?layout=[MapPanel,DistBarChart]-[MapPanel,ScatterPlot]-[MapPanel,BarChart]-[MapPanel,DataTable]&flashLayout=(2,2,2,2)&flashComponents=(MapPanel,DistBarChart),(MapPanel,ScatterPlot),(MapPanel,BarChart),(MapPanel,DataTable)&componentIds=(map,groupBarPanel),(map,scatterPlot),(map,barPanel),(map,)&enableAnimation=true,true,true,true&story=https://ec.europa.eu/dashboard/dashboard/18/1545.xml&explorerURL=https://ec.europa.eu/dashboard/group/9/&regionSelectorVisible=false&showRecordSelector=closed&showSnapshotButtons=false&showMetaData=true&metaOnTop=true&showTitle=false&metaHeight=30&colorLegendVisible=auto&showEmbedButton=false&linkIndicators=true&linkAxis=x&mode=html5&fallback=true&mapIgnore=true&locale=en](https://ec.europa.eu/dashboard/publisher/vislet.html?layout=%5bMapPanel,DistBarChart%5d-%5bMapPanel,ScatterPlot%5d-%5bMapPanel,BarChart%5d-%5bMapPanel,DataTable%5d&flashLayout=(2,2,2,2)&flashComponents=(MapPanel,DistBarChart),(MapPanel,ScatterPlot),(MapPanel,BarChart),(MapPanel,DataTable)&componentIds=(map,groupBarPanel),(map,scatterPlot),(map,barPanel),(map,)&enableAnimation=true,true,true,true&story=https://ec.europa.eu/dashboard/dashboard/18/1545.xml&explorerURL=https://ec.europa.eu/dashboard/group/9/&regionSelectorVisible=false&showRecordSelector=closed&showSnapshotButtons=false&showMetaData=true&metaOnTop=true&showTitle=false&metaHeight=30&colorLegendVisible=auto&showEmbedButton=false&linkIndicators=true&linkAxis=x&mode=html5&fallback=true&mapIgnore=true&locale=en)

The dataset contains the following variables:

[1] "nuts2" "country" "idcountry" "idregion"

[5] "region" "ojv" "m\_ojv" "gdp\_cmp"

[9] "gdp\_ha" "gva" "gva\_growth" "income\_hab"

[13] "purchase\_power\_hab" "empl\_rate\_f" "empl\_rate\_m" "empl\_rate\_tot"

[17] "unempl\_rate\_f" "unempl\_rate\_h" "unempl\_rate\_tot" "pop\_tot"

[21] "pop\_density"

With ;

m\_ojv: CEDEFOP OJVs (monthly average)

gdp\_cmp: GDP at current market prices

gdp\_ha: : GDP at current market prices per inhabitant

gva: gross value added

empl, unempl, : rates of employment, unemployment

In total, the dataset contains 326 NUTS2 regions for 28 countries. Ireland with missing data for GDP was not retained in the modelling exercise. We therefore have a set of 210 regions for 12 countries in the modelling set and a prediction set of 114 regions for 19 countries.

After a few trials, the following linear model with 3 variables (population, GDP and rate of employment) was selected:

lm(formula = log(m\_ojv) ~ log(pop\_tot) + log(gdp\_cmp) + empl\_rate\_tot,

data = inp1)

with the following results for the regression:

Residuals:

Min 1Q Median 3Q Max

-2.83495 -0.45271 -0.08918 0.48192 2.63985

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -16.319066 1.690733 -9.652 < 2e-16 \*\*\*

log(pop\_tot) 1.074106 0.060450 17.769 < 2e-16 \*\*\*

log(gdp\_cmp) 0.724547 0.168250 4.306 2.57e-05 \*\*\*

empl\_rate\_tot 0.040354 0.006669 6.051 6.71e-09 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.7007 on 206 degrees of freedom

Multiple R-squared: 0.6923, Adjusted R-squared: 0.6879

F-statistic: 154.5 on 3 and 206 DF, p-value: < 2.2e-16

The scatterplots of OJVs againts population and GDP at the NUTS2 level are shown on Figure 13 and

Figure 14. Some outliers appear on these plots such as the Paris region for population and London for the GDP.

Figure 13 Scatterplot of NUTS2 CEDEFOF OJVS (monthly average) against population size

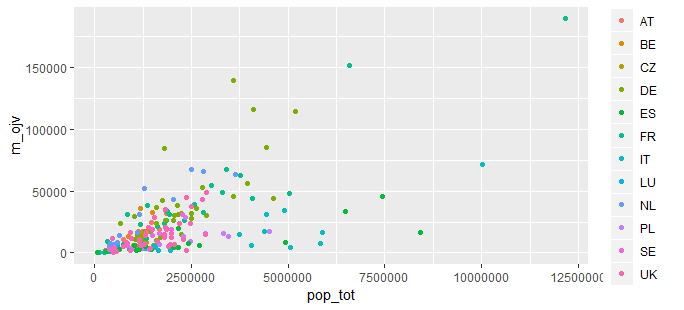
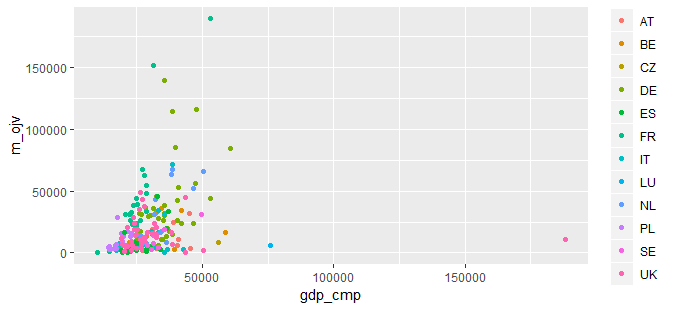
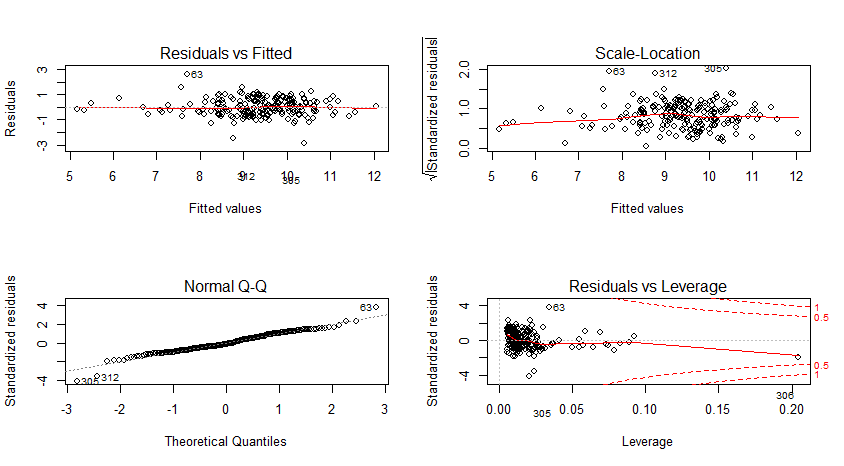


Figure 14 Scatterplot of NUTS2 CEDEFOF OJVS (monthly average) against GDP



**Residuals**: apart from these few outliers, plots of residuals (Figure 15) do not exhibit particular features which would need additional treatments.

Figure 15 Standard plots of residuals



**Fit and predictions :** Fit of model and predictions for countreis wit hand without CEDEFOF data are shown on Figure 16 and Figure 17.

Figure 16 Scatterplot of fitted OJVs against observed values (log scale; corresponds to Figure 6).

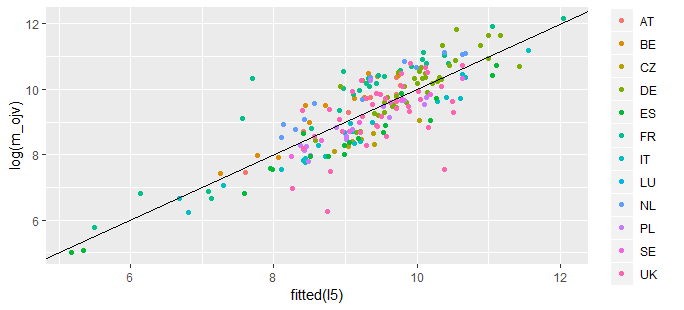
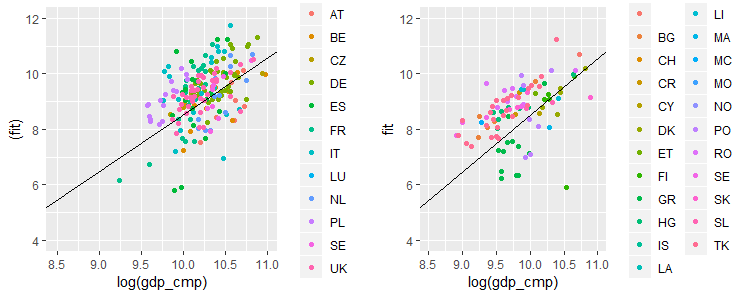


Figure 17 Scatterplot of predicted OJVs against GDP (log scale; corresponds to Figure 7).

1. NUTS2 regions in CEDEFOP dataset b) remaining NUTS2 regions



Conclusions :

Similar conclusions can be drawn from analysis of the data at the regional level as compared to the national level. The model could be improved using a robustification approach to handle outliers.

### Further improvements (Sophie in work)

1. Replace population size by employment (jobs, occupied posts or proxies such as employees (STS220), or persons in employment from LFS)
2. Produce a model with intercept = 0 (a country with pop and GDP = = 0 would have 0 OJVs)
3. Introduce robustification in the model in order to handle outliers (cf regional analyses below)

### Further improvements (Sophie in work)

1. Replace population size by employment (jobs, occupied posts or proxies such as employees (STS220), or persons in employment from LFS)
2. Produce a model with intercept = 0 (a country with pop and GDP = = 0 would have 0 OJVs)
3. Introduce robustification in the model in order to handle outliers (cf regional analyses below)
4. Visual reprentation (regional maps)

# Overall summary of the study of the relationships between OJVs and socio-economical variables

To shortly summarize the two preceding chapters, CEDEFOP data investigated here clearly demonstrate a srong relationship between OJVs and socio-economical variables such as national and regional population sizes and GDPs. These relationships can be modelled with promising outputs. Such models may potentially be used for predictive purposes. For instance, since OJVs are portentially instantly available, they might offer a data source for developing Flash estimates of national and regional GDPs . The models and choice of variables should of course be studied in more detail and improved. The models also indicate that the offer the potential of predicting OJVs for countries without data for OJVs on the basis of their socio-economical variables. Such cases without CEDFOP OJVs data, offer an interesting opportunity to test the models (blind approach).

# Comparisons between OJVs and Job vacancy Statistics (JVS)

A major issue when dealing with OJVs is to understand how Online Job advertisements are connected to actual job vacancies, in particular to the official Job Vacancy Statistics. For this purpose, CEDEFOP data are in a first step compared to the Eurostat JVS downloaded from :<https://ec.europa.eu/eurostat/databrowser/view/tps00172/default/table?lang=en>.

Rescaling the data: In an attempt to bring the data on a common basis allowing to compare the two sources, the total amount of CEDEFOP OJVs (a cumulated sum over a 9 months period : July2018-March2019) are transformed to monthy averages, as follows: m\_ojv=total CEDEFOP OJVs/9. On the contrary, the quarterly Eurostat JVS correspond to a snapshot taken at the reference day (the last day of the quarter). For the same period, we have data of three quarters (2018Q3, 2018Q4 and 2019Q3). The absolute amount of JVS are averaged as follows: m\_jvs= (JVS183 + JVS184 + JVS191)/3.

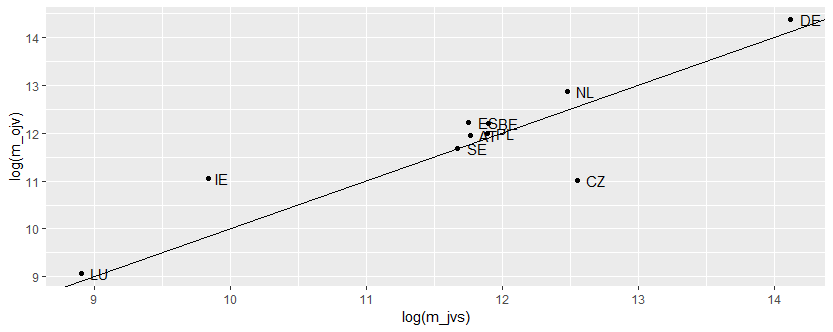
Table 4 presents the CEDEFOP OJVs (9-months averages of total number of OJVs for the period July 2008 to March 2019) in comparison with absolute figures of the official published JVS of Eurostat (3-quarters averages) for the 13 countries with CEDEFOP data. Note that no data are not available for two large countries (FR and IT). We are therefore left with 11 cases for comparisons. Apart for CZ and IE for which figures diverge largely, monthly OJVs are of the same order of magnitude with OJS exceeding JVS from 10 to 61% (Table 4; Figure 18). Note the very close agreement between the two indicators for Sweden.

Table 4 Table of averaged CEDEFOP OJVs and Eurostat JVS figures.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | Eurostat | CEDEFOP | ratio |
| country | idcountry | m\_jvs | m\_ojv | r\_ojv\_jvs |
| Poland | PL | 145'547.70 | 160'476.70 | 110% |
| Italy | IT | NA | 263'448.30 | NA |
| Spain | ES | 126'683.30 | 203'600.00 | 161% |
| Czechia | CZ | 281'311.00 | 60'216.30 | 21% |
| Luxembourg | LU | 7'364.70 | 8'617.70 | 117% |
| Sweden | SE | 116'259.70 | 118'166.40 | 102% |
| Ireland | IE | 18'700.00 | 63'340.00 | 339% |
| Belgium | BE | 146'202.70 | 200'265.60 | 137% |
| Austria | AT | 128'471.70 | 157'087.10 | 122% |
| France | FR | NA | 1'328'799.60 | NA |
| Germany | DE | 1'346'804.00 | 1'742'199.20 | 129% |
| United Kingdom | UK | 846333.3 | 1'478'181.80 | 175% |
| Netherlands | NL | 261'577.30 | 387'097.20 | 148% |

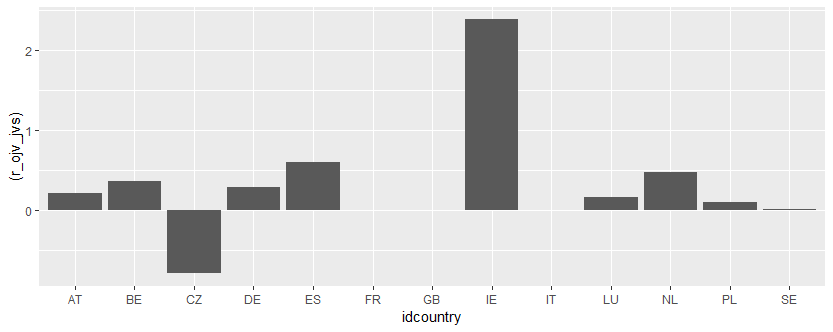
Figure 18 Graphical representation of data of Table 4 (log scales).

The abline of slope 1 makes conspicuous two oulierrs (IE and CT) for which OJVs depart largely from the Eurostat JVS



The differences to a 1/1 ratio OJVS/ JVS is also represnted graphically on Figure 19.

Figure 19 Graphical representation of the fidtance to a 1/1ratio of OJVS/ JVS with extreme values in opposite directions for IE and CZ.



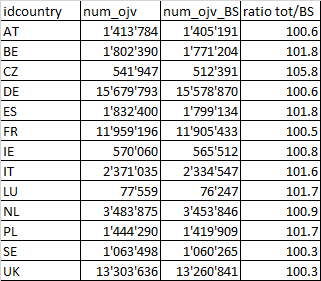
Improving the comparability of the two data sources

The comparability of the two data sources may be improved in several ways;

1. Subsetting the OJVs to the corresponding months for which JVS results are available, i.e. September and December 2018 and March 2019
2. Subsetting the OJVs data to job advertisements still open for the JVS reference days, i.e. last working days of September and December 2018 and March 2019
3. Subsetting the OJVS to job advertisements corresponding to the NACE2 categories covered by the official Eurostat JVS (sections B-C), or to the NACE2 branches actually covered for each country individually
4. Comparing the monthly OJVs (as above subsetted) data with the corresponding month of the quarterly JVS

Subset of CEDEFOP data for NACE B to S: the subset of CEDEFOP data for NACE B to S indicates that OJVs for NACE A, T, U represent only a minute part of total collected OJVs and may only explain a small part of the differences between OJVs and JVS. In this case of CZ, it would even increase the gap between the two data sources.

Table 6 Subset of NACE\_BS relatively to total OJVs per country.



Monthly vs quarterly data:

Although the results of Table 4 are based on rough estimates (averages of total numbers of OJVs), they suggest a consistent relationship between OJVs and JVS and leave some hope that JVS may be approached using OJVS. Table 5 provide preliminary results supporting this hypothesis for the 11 countries for which CEDEFOP and JVS data are available. In the first three columns official Eurostatm JVS figures are reported for quarters 2018Q3, 2018Q4 and 2019Q1. A selection of OJVs with similar orders of magnitude is given in the next three columns, and ratios in % are given in the last three columns. OJVs have been selected as follows:

1. for each quarter, we selected the OJVs open during a window of 30 days corresponding to the last month of the quarter which were still open at the reference day have been counted (i.e 1-30. September 2018, 1-31 December 2018 and 1-31 March 2019 using the variables grab\_date and expire\_date)
2. for each country we retained only the OJVs corresponding to a given country and advertised in this country
3. we only retained OJVs for NACE B to S
4. we excluded advertisements from sources coming from an agency abroad the country (for instance a job advertisement from a source like CZ\_GIGAJOB in Czech Republic for a job in Ireland was rejected)
5. when the OJVs figures differed too much from the JVS, a closer agreement was searched playing with the time window

Table 5 Attempt to extract monthly OJVs coresponding to quarterly JVS for period 2018Q3 to 2019Q1.



Results of Table 5 indicate that a reasonable agreement can be reached in most cases with a 30-days time window corresponding to the last month of the quarter. In the case of Ireland the window was reduced to 15 days, while it has to be increased to 2 or 3 months for LU and BE. A closer analysis of the data indicates that the raw data may be very heterogenous within a source, with missing data or large numbers of collected OJVs among months within the same source. Much work is still needed to better understand the data at this level of detail, as well as to validate/invalidate the approach of data selection attempted here.

The case of Ireland: detailed OJVs data for Ireland illustrate the heterogeneity of the OJVs data. (to be continued)

|  |  |
| --- | --- |
| 1. OJVs grabed monthly for Ireland | 1. Same data after selection for NACE BtoS and exclusion of sources from abroad |
|  |  |
|  | |

# Understanding the OJVs business models

This issue also enhances an important point. Online job portals have not been developed in order to produce statistics or indicators of the job market. Nor have they been invented for altruistic purposes. Their goal is different. They have been developed on economical grounds and for business reasons, crudely said “to make money” in a way or another. Therefore, a good understanding of the “business models” underlying the development of job portals, as well as of the market of online job advertisements is crucial. A good understanding of their mechanisms is a key issue to understand the data and find convincing ways to use them for statistical purposes.

# Annexes:

## Detailed analysis of particular cases and extremely large OJVs

The graphs and tables below give some examples of scatter of regional OJVs data within countries with respect to GDP. In most cases, a few extremely large values appear in the data. In most cases, large figures correspond to leading parts of the countries, such as Paris or Berlin. In Italy, large numbers of OJVs appear in the northern regions, while Catalunia is leading before Madrid in Spain. Inner London West offers another kind of outlier with an extremely large GDP per inhabitant with respect to other regions in the UK(Figure 21 ).

Figure 20 Scatter plots of OJVs versus GDP for Italy

Table 7 Italy: NUTS2 OJVs and GDP ordered by descending OJVS showing the regional differences between northern and other NUTS2 regions

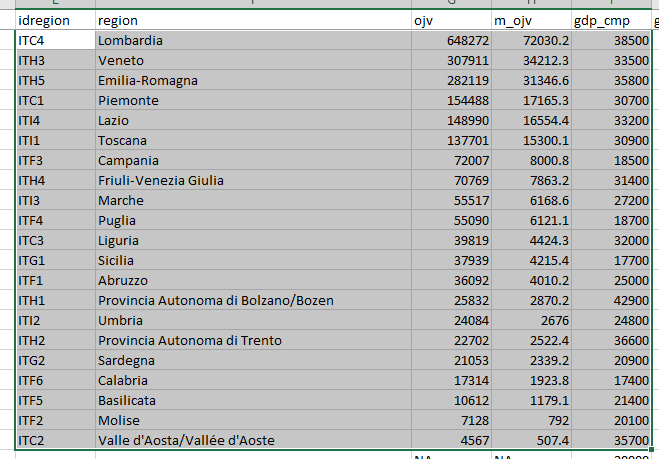


Figure 21 Scatter plots of OJVs (vertical axis) versus GDP for Germany

Table 8: NUTS2 OJVs and GDP ordered by descending OJVS

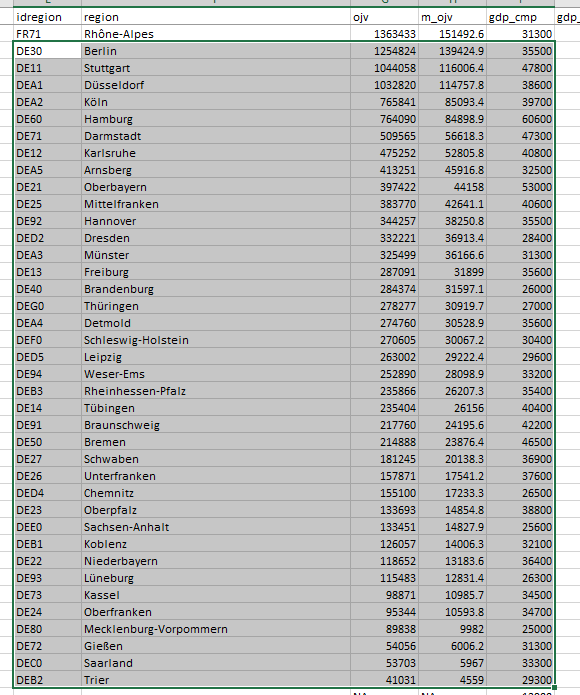


Figure 22 Scatter plots of OJVs (vertical axis) versus GDP for Spain

Table 9 Spain: NUTS2 OJVs and GDP ordered by descending OJVS

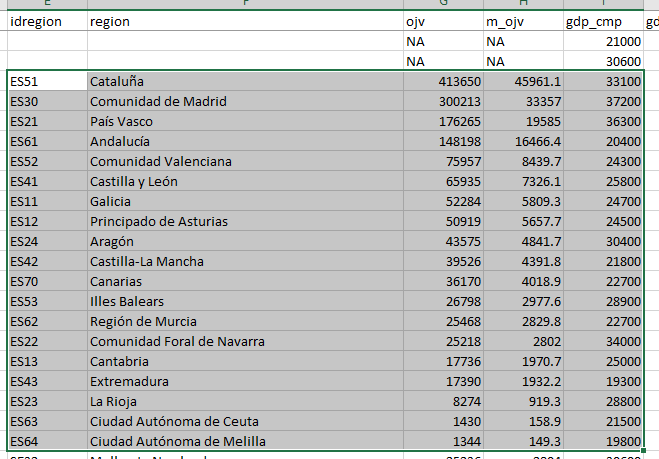
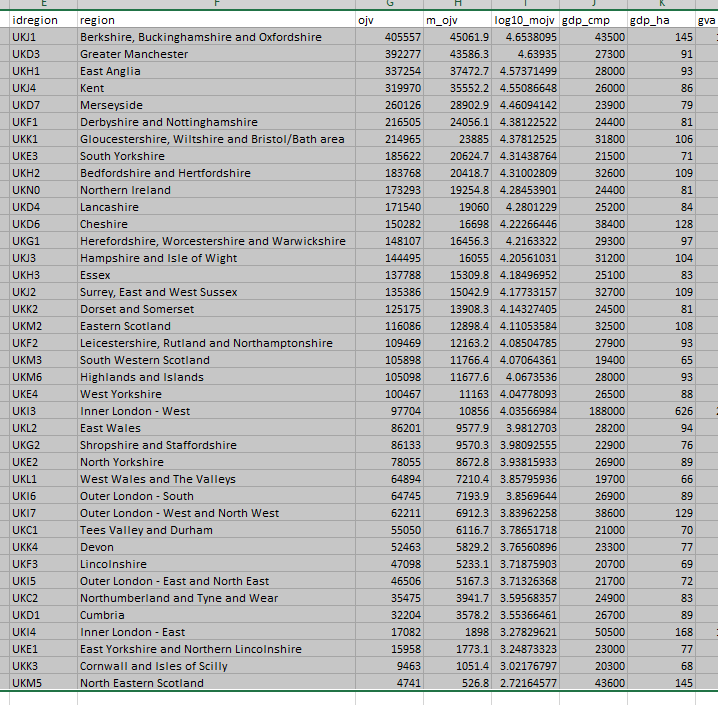


Figure 23 Scatter plots of average monthly OJVs(vertical axis) versus GDP for the United Kingdom.

Note the extremely large GDP for DE94 (Inner London – West)

Table 10 NUTS2 OJVs and GDP ordered by descending OJVS



## Problems encountered in CEDEFOP Data

Variable idcountry: should be adapted to official abbreviations

|  |  |
| --- | --- |
| (preliminary list for validation)  short long | |
| RU | Russian Federation |
| NO | Norway |
| FR | France |
| SE | Sweden |
| BY | Belarus |
| UA | Ukraine |
| PL | Poland |
| AT | Austria |
| HU | Hungary |
| MD | Moldova |
| RO | Romania |
| LT | Lithuania |
| LV | Latvia |
| EE | Estonia |
| DE | Germany |
| BG | Bulgaria |
| GR | Greece |
| AL | Albania |
| HR | Croatia |
| CH | Switzerland |
| LU | Luxembourg |
| BE | Belgium |
| NL | Netherlands |
| PT | Portugal |
| ES | Spain |
| IE | Ireland |
| IT | Italy |
| DK | Denmark |
| UK | United Kingdom |
| IS | Iceland |
| SI | Slovenia |
| FI | Finland |
| SK | Slovakia |
| CZ | Czech Republic |
| BA | Bosnia and Herzegovina |
| MK | Macedonia |
| RS | Serbia |
| ME | Montenegro |
| XK | Kosovo |