

**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 time series**

**using the 3rdCEDEFOP Dataset**

**(March 2020)**

**Prepared by:**

**Swiss Federal Statistical Office**

**(Francis Saucy)**

**Date of present version: 26th March 2020**

Workpackage Leader:

Tomaž Špeh (SURS, SI)

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

mobile phone: +38651672116

Contents

[Analyses of time series using the 2nd CEDEFOP Dataset 3](#_Toc36058861)

[Goals of the report 3](#_Toc36058862)

[The main goals of this report are to: 3](#_Toc36058863)

[a) describe the OJA CEDEFOP data in terms of time series 3](#_Toc36058864)

[b) compare OJA CEDEFOP series with the Eurostat JVS time series 3](#_Toc36058865)

[Technical considerations 3](#_Toc36058866)

[Data description 3](#_Toc36058867)

[Monthly time series 4](#_Toc36058868)

[Quartely OJA time series 6](#_Toc36058869)

[Conclusions: 10](#_Toc36058870)

**List of Figures**

Figure 1 Graphical representation of monthly OJA from Table 1 (absolute figures V2 before 25.03.2020; 15 months) 5

Figure 2 Graphical representation of data from Figure 1 and Table 1 rescaled as differences relative to the mean (in percent; V2 before 25.03.2020; 15 months) 5

Figure 3 Graphical representation of data from Table 2 : Extract of unique OJA by country and grab\_month rescaled as differences relative to the mean (in percent; V3: 25.03.2020; 18 months) 5

Figure 4 Ouarterly OJA time series from the CEDEFOP dataset for the 28 countries (period 2018Q3 to 2018Q3; relative variations to the average in %; V2 before 25.03.2020; 5 quarters). 7

Figure 5 Ouarterly OJA time series from the CEDEFOP dataset for the 28 countries (period 2018Q3 to 2018Q3; relative variations to the average in %; V3 25.03.2020; 6 quarters). 7

Figure 6 Quartely fluctuations of official Eurostat JVS relative to the mean for 18 EU countries (NACE sections A-S; period 2016-2019) 8

Figure 7 Quartely fluctuations of official Eurostat JVS relative to the mean for the period 2018Q3 to 2019Q4. Total number of vacancies, sections A-S of the NACE Rev.2., 18 countries. Source : https://ec.europa.eu/eurostat/web/labour-market/job-vacancies/database file:jvs\_q\_nace2.tsv. 8

**List of Tables**

[Table 1 Extract of unique OJA by country and grab\_month (V2 before 25.03.2020; 15 months) 4](#_Toc36058884)

[Table 2 Extract of unique OJA by country and grab\_month (V3: 25.03.2020; 18 months) 4](#_Toc36058885)

[Table 3 Compared volatility between CEDFOP OJA (3rd dataset) and official JVS for period 2018Q3 to 2019 Q4. Difference relative to the average in %) 9](#_Toc36058886)

# Analyses of time series using the 2nd CEDEFOP Dataset

# Goals of the report

# The main goals of this report are to:

# describe the OJA CEDEFOP data in terms of time series

# compare OJA CEDEFOP series with the Eurostat JVS time series

1. undertstand causes of divergences

# Technical considerations

The analyses have been conducted during the second half of March 2020 (18th to 30th). They refer to the set of V2 CEDEFOP data version provided in the beginning of 2020. They cover the period July 2018- September 2019, i.e. a period of 15 months or 5 quarters for all 28 EU countries. Data have been accessed using Jupyter notebooks on the new BDTI platform (cloud amazonaws). Statistical analyses have been conducted using the open source software R.

# Data description

The V2 dataset encompasses 107.83 millions of records (V1: 64.3; +43.5 millions; +67.6%). After exclusion of duplicate cases (query: SELECT COUNTRY, COUNT(DISTINCT GENERAL\_ID) as num\_job\_vacancy), the V2 dataset displays 92.4 millions of unique records, or “OJAs = Online job advertisements” (V1: 55.5 millions; +36.9 millions; + 66.5%).

From 25th of March 2020 a 3rd version of the CEDEFOP dataset is apparently online, with dramatically less OJA than in the second dataset, i.e. 85 instead of 108 millions and 67.5 instead of 92.4 millions of unique records (-25 millions). In all countries the global number of OJA is drastically lower. Moreover, the remaining data are dispatched over 18 months (July 2018 Dec. to 2019) instead of 15. Which means that data of the 3rd dataset cover 6 quarters, instead of 5.

# Monthly time series

Table 1 Extract of unique OJA by country and grab\_month (V2 before 25.03.2020; 15 months)

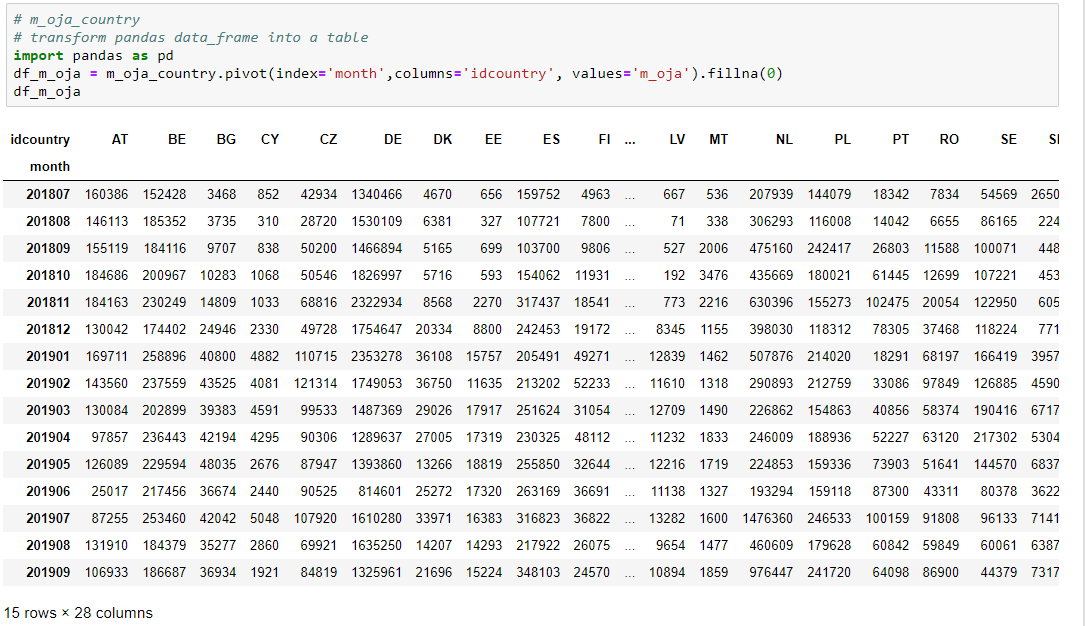


Table 2 Extract of unique OJA by country and grab\_month (V3: 25.03.2020; 18 months)

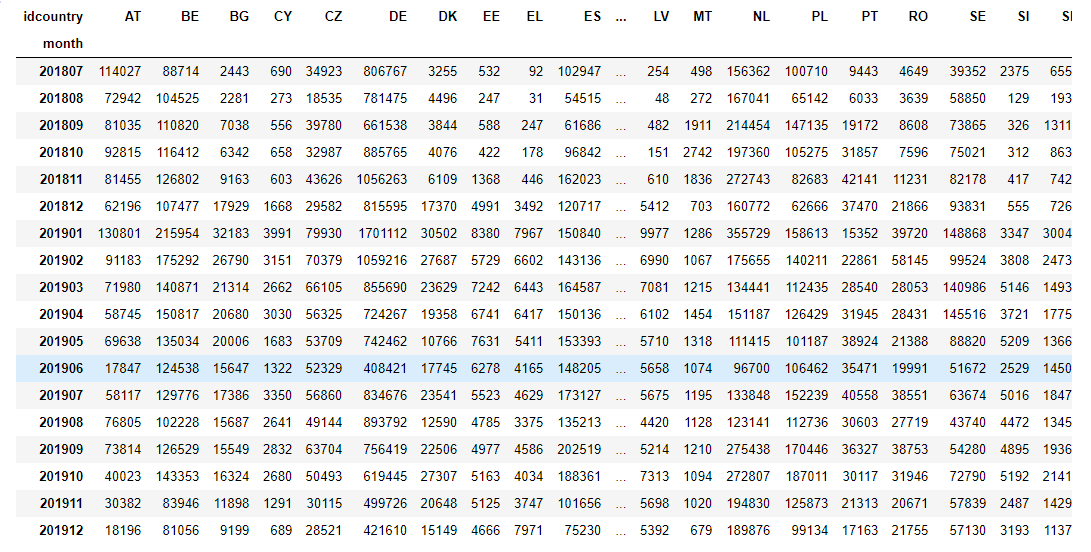


Figure 1 Graphical representation of monthly OJA from Table 1 (absolute figures V2 before 25.03.2020; 15 months)



Figure Graphical representation of monthly OJA from Table 2 (absolute figures V3 :25.03.2020; 18 months)

Time series display drastic modifications in their patterns on 25.03.2020 as compared to earlier data extracts for many countries (e.g AT, NL, UK, etc).

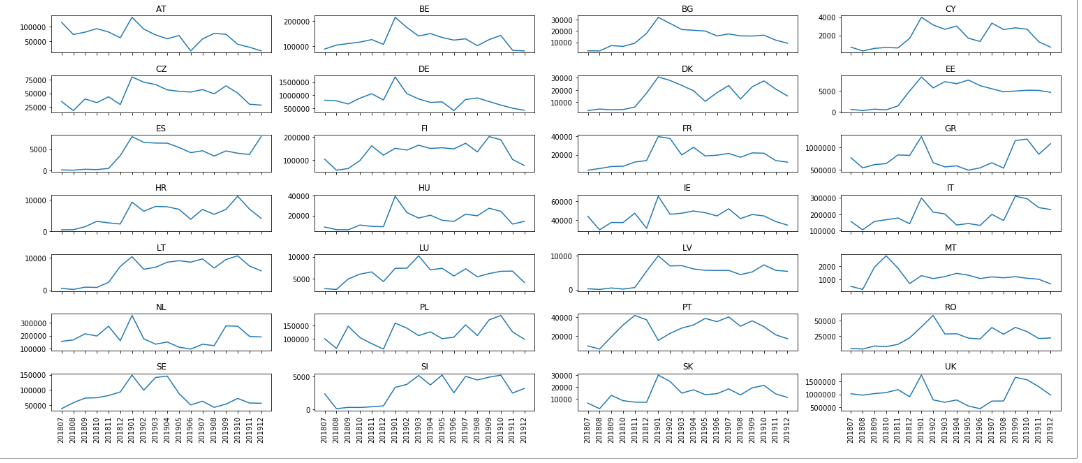


Figure 3 Graphical representation of data from Figure 1 and Table 1 rescaled as differences relative to the mean (in percent; V2 before 25.03.2020; 15 months)

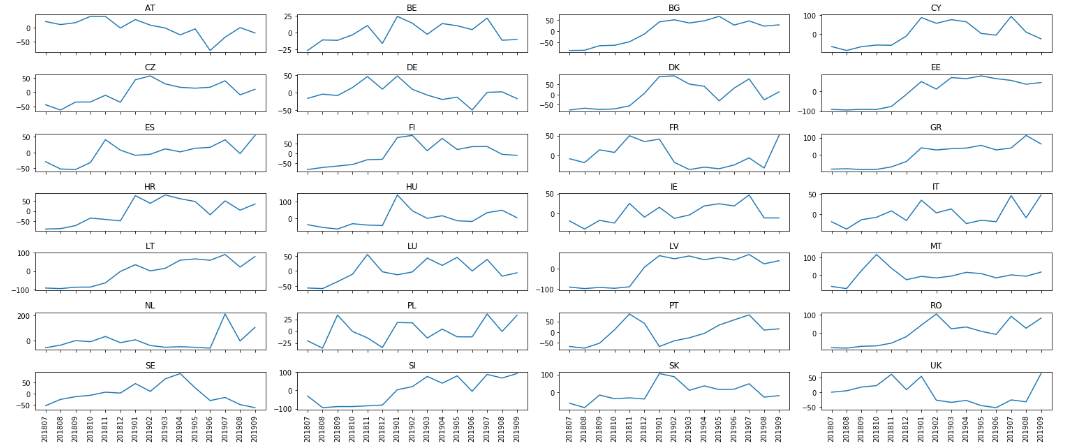
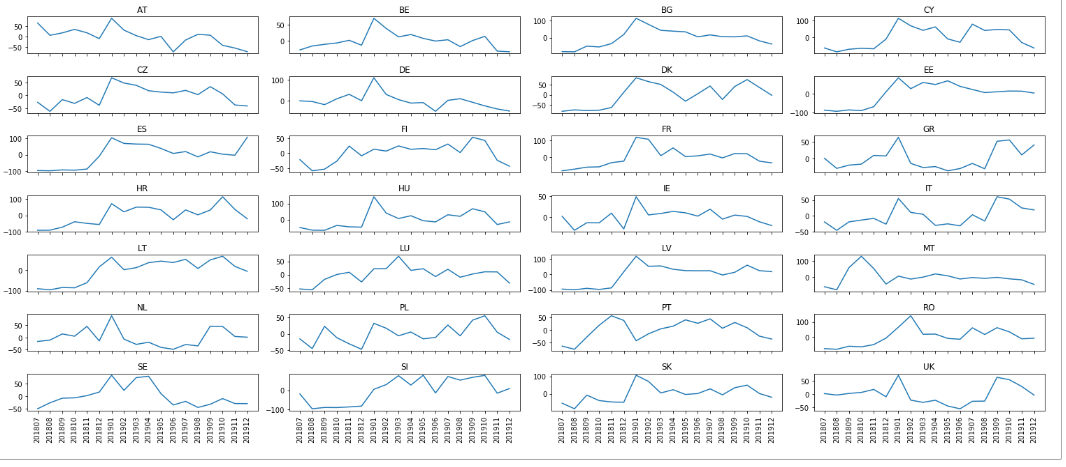


Figure 4 Graphical representation of data from Table 2 : Extract of unique OJA by country and grab\_month rescaled as differences relative to the mean (in percent; V3: 25.03.2020; 18 months)



Monthly OJA time series, defined by the month of actual scraping, are displayed in absolute figures in Table 1, Table 2, Figure 1, Figure 2 and Figure 3. These data show a wide volatility among months. In order to appreciate this variability, the same data have been rescaled in Figure 2 and in Figure 3 in relative figures as differences to the average (expressed in percentage). In most cases, the differences to the mean vary within a range of at least 50% (e.g. -25 to 25% for BE), but more often in a range of 100 or more

**Patterns of fluctuations**: the following possible patterns appear at visual inspection of the graphs:

**2nd dataset** (period July 2018 - Sept. 2019; data extracted before 25.03.2020)

1. lower numbers in 2018 than in 2019 : BG, CY, CZ, DK, EE, FI, GR, HR, HU?, LT, LV, RO, SI, SK, i.e. mostly countries for which no data were available in the first CEDEFOF dataset
2. lower numbers in 2019 than in 2018 : AT, DE, FR and UK, i.e. countries for which data were available in the first CEDEFOF dataset
3. no clear difference among months for the remaining countries: i.e., BE, ES, IE, IT, , PL, SE, mostly countries for which data were available in the first CEDEFOF dataset
4. little fluctuations with few major outliers : NL (+200%) in July 2017 and MT (+100%, oct. 2018), HU?
5. seasonal pattern: PT?

**3rd dataset** (period July 2018 - Dec. 2019; data extracted on 25.03.2020)

The patterns described for the 2nd dataset as well at the volatility are stongly attenuated in the 3rd version. See for instance, the outlier of NL in the 2nd dataset disappears and the series becomes very different the 3rd dataset.

This attempt of classification is tentative and one could order cases differently. It nevertheless suggests a large heterogeneity of OJA fluctuations among countries and months. These patterns as well as the differences between the 2nd and 3rd datasets need to be explained and are worth further investigations in order to better understand the significance of the data. Several possible causes will be addressed.

# Quartely OJA time series

In order to compare these results with official Eurostat Job vacancy statistics (JVS), a common framework of comparison is necessary. JVS are only available on a quarterly basis and for a selection of branches of activities according to the NACE. The largest scope for JVS covers the sections A-S. JVS time series for the period covered by the CEDEFOP second dataset are available for 18 EU countries at the A-S level of aggregation.

In a first step, quarterly time series have been drawn from the CEDEFOP dataset selecting data for the months corresponding to the last month of each quarter, March, June, September and December, i.e. 5 quarters for the 2nd version (from Sept. 2018 (2018Q3) to Sept. 2019 (2019Q3), and 6 quarters for the 3rd version. These time series are displayed on Figure 4.

They also exhibit a large volatility, in the range of 50 to 100% for most cases.

Figure 5 Ouarterly OJA time series from the CEDEFOP dataset for the 28 countries (period 2018Q3 to 2018Q3; relative variations to the average in %; V2 before 25.03.2020; 5 quarters).

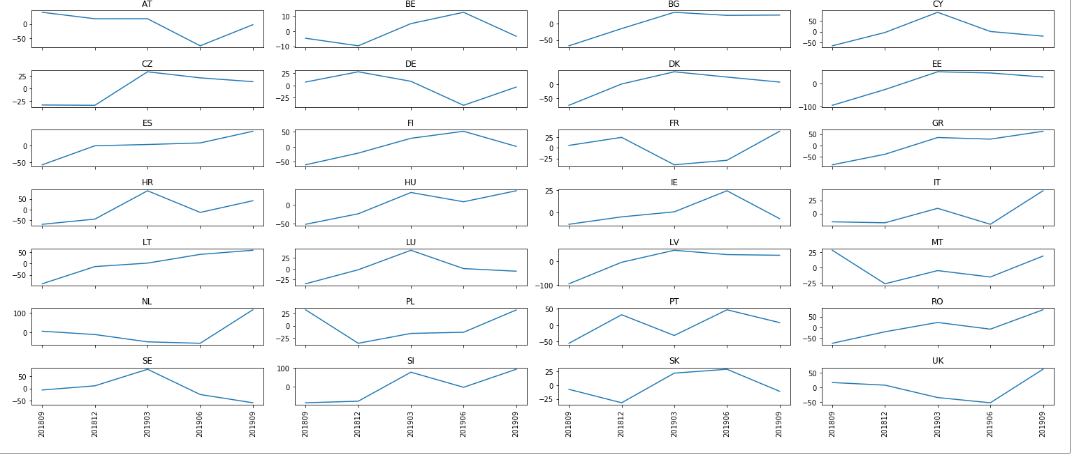
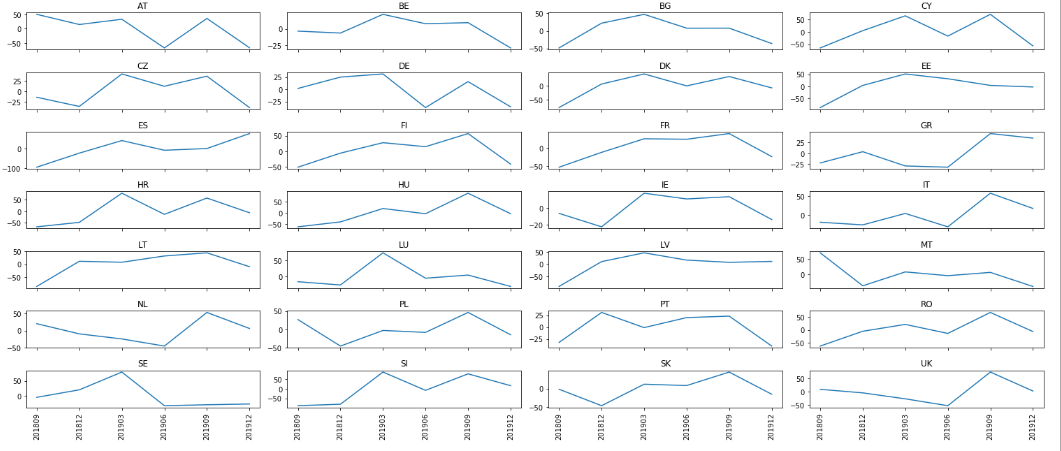


Figure 6 Ouarterly OJA time series from the CEDEFOP dataset for the 28 countries (period 2018Q3 to 2018Q3; relative variations to the average in %; V3 25.03.2020; 6 quarters).



**Patterns of fluctuations of EU JVS**

EU JVS fluctuations (Figure 6 and Figure 7) display a much reduced volatility , as compared to CEDEFOP OJA (Figure 4 and Figure 5). The difference to the mean of the series is on the average of 7.0% for JVS (Min: 2.7, Max: 15.4) as compared to 30.8% (Min: 13.2, Max: 59.2) for OJA (Table 4).

A clear seasonality appears for most countries in the four-year (2016-2019) time series (12 from 18, i.e. BG, EE, HR, HU, LU, LV, NL, NO, PL, PT, SE, SI; Figure 6) with highest values from spring to autumn and lowest during winter. No such pattern appears in the shorter quarterly time series (Figure 7).

Figure 7 Quartely fluctuations of official Eurostat JVS relative to the mean for 18 EU countries (NACE sections A-S; period 2016-2019)

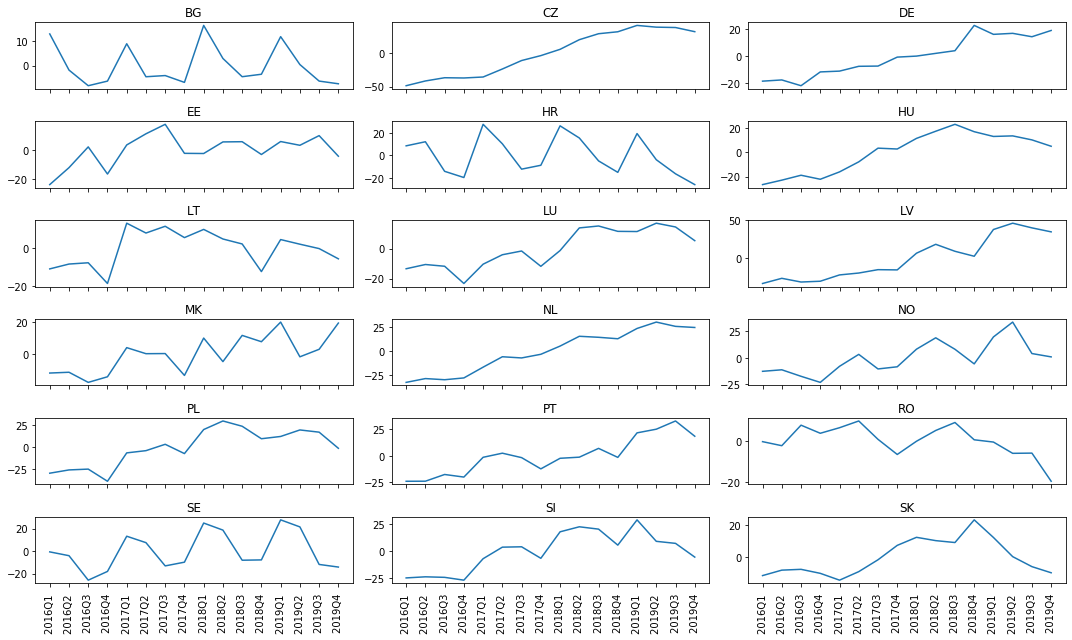


Figure 8 Quartely fluctuations of official Eurostat JVS relative to the mean for the period 2018Q3 to 2019Q4. Total number of vacancies, sections A-S of the NACE Rev.2., 18 countries. Source : <https://ec.europa.eu/eurostat/web/labour-market/job-vacancies/database> file:jvs\_q\_nace2.tsv.

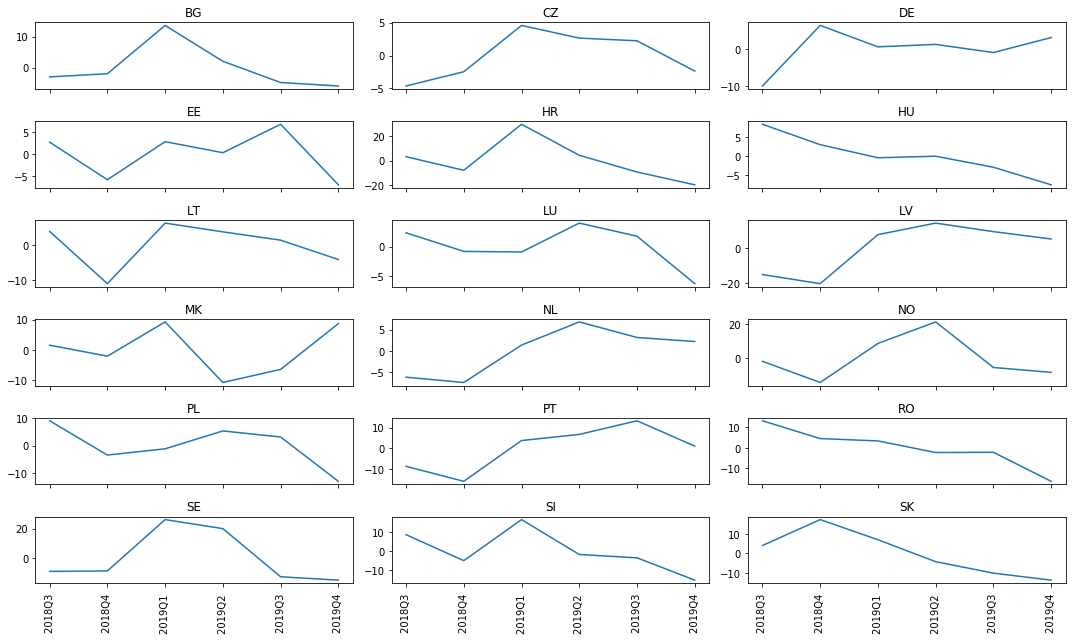


Table Comparisons between CEDEFOP OJA (3rd dataset) and official JVS for period 2018Q3 to 2019 Q4. Absolute figures.



Table 4 Compared volatility between CEDEFOP OJA (3rd dataset) and official JVS for period 2018Q3 to 2019 Q4. Difference relative to the average in %)

