



International mobility between the UK and Europe around Brexit: a data-driven study

Alina Sirbu¹ · Diletta Goglia^{1,2} · Jisu Kim³ · Paul Maximilian Magos¹ · Laura Pollacci¹ · Spyridon Spyratos⁴ · Giulio Rossetti⁵ · Stefano Maria Iacus⁶

Received: 20 September 2023 / Accepted: 26 March 2024 / Published online: 20 April 2024
© The Author(s) 2024

Abstract

Among the multiple effects of Brexit, changes in migration and mobility across Europe were expected. Several studies have analysed these aspects, mostly from the point of view of perceptions, motivations, economic effects, scenarios, and changes in migration from Central and Eastern European countries. In this study we propose an analysis of migration and cross-border mobility using an integrated data-driven approach. We investigate official statistics from Eurostat, together with non-traditional data, to give a more complete view of the changes after Brexit, at EU and regional level. Specifically, we employ scientific publication and Crunchbase data to study highly-skilled migration, Twitter and Air Passenger data to investigate monthly trends. While main trends are preserved across datasets, with a general decrease in migration towards the UK immediately after the referendum approval, we are able to also observe more fine grained trends specific to some data or regions. Furthermore, we relate the changes in mobility observed from Air Passenger data with attention to Brexit from Google Trends data.

Keywords Human migration · Human mobility · Brexit · Eurostat · Air passenger data · Twitter data · Scientific publications · Crunchbase · Google Trends

Introduction

International mobility includes various mobility types, such as short-term travelling for work or tourism, medium-term or seasonal migration, and long-term migration. Mobility is determined by a plethora of socio-economic, political, environmental,

Google Scholar: [D.G.](#), [L.P.](#), [A.S.](#)

[J.K.](#)

[S.S.](#)

[G.R.](#)

[S.M.I.](#)

European Union Intellectual Property Office: views and opinions set out in this article are strictly personal and should not be attributed to the EUIPO.

Extended author information available on the last page of the article

health and individual factors, and the relationship between them has been long studied [23]. In particular, changes in the geopolitical organisation of the world may be very disruptive for mobility in general and migration in particular. In this work we investigate, using a data-driven approach with multiple data sources, the effects of Brexit on mobility between the United Kingdom (UK) and the European Union (EU27).

The UK formally withdrew from the European Union (EU) on the 31st of January 2020. On the 23rd of June 2016, the UK held a referendum on its membership in the European Union. The UK citizens voted to leave the EU by 52%–48%.¹ Starting from the referendum results, the UK Government formally notified the European Council of the decision to trigger Article 50² of the Lisbon Treaty on the 29th of March 2017. This date signs the actual start of the Brexit process,³ when a two-year countdown began to the formal leaving of the UK from the EU. Several draft agreements were reached and rejected afterwards, with deadlines being postponed several times (from the 29th of March 2019 to the 30th of June 2019, then to the 31st of October 2019) until the 31st of January 2020. At 11 pm on the 31st of January 2020, the UK left the EU, entering a transition period. The transition ended at 11 pm on the 31st of December 2020 when the United Kingdom also left the EU single market.

In the months before the referendum and then afterwards until the actual exit from the EU, there has been increased media attention on the process, both in the UK and Europe. Despite immigration being central to the politics of Brexit, the pre-referendum discussion was dominated by its economic consequences [38, 39]. Since the referendum, immigration has become a less salient political issue, and public attitudes toward immigration seem to have become more positive [38]. However, migration flows from the EU fell sharply, attracting more attention to its economic implications. After the expansion of the EU (2004, 2007, and 2011), the movement of people between the UK and EU countries grew notably. When the referendum took place in June 2016, net migration to the UK achieved a record. Since the first months after the referendum, post-Brexit scenarios involving economic implications and labour migration have been speculated. These discussions concern the uncertainty and vulnerability that both the transition period and the after-Brexit period caused to foreign workers in the UK and to UK citizens residing in the EU Member States [4]. Studies have been produced in an attempt to analyze and forecast many possible outcomes from different perspectives, from political and legal to economic and social. Less than a year from the referendum, [43] already stated that inflows of foreign researchers and academics in the UK would have suffered from the implication of Brexit, and even earlier [50] was assuming lower levels of immigration and higher levels of emigration in the country due to a post-Brexit economic decline of

¹ <https://eur-lex.europa.eu/content/news/Brexit-UK-withdrawal-from-the-eu.html>.

² Article 50 is the only legal mechanism for a member state of the European Union (EU) to leave the Union. The article describes the steps a member state must go through to withdraw from its obligations (Source: <https://commonslibrary.parliament.uk/research-briefings/cbp-7960/>).

³ Note that the referendum is legally non-binding for the purposes of Article 50.

the UK. More recent studies have focused their attention on the consequences of Brexit on highly-skilled migration [10, 29, 47] and on the significant fall of migration inflows and outflows in the UK after June 2016 [38].

This paper contributes to the analysis of the effects of Brexit by studying mobility between the UK and EU27 in the last decades. The objective is to obtain a global view of mobility trends—from different viewpoints—and to understand whether and how Brexit has affected mobility. The analysis is based on several datasets covering different types of mobility and demographic groups for an integrated quantitative image. The first data type that we investigate is migration flows from the Eurostat database. These give official information on long-term migration, on a yearly basis, and are traditional data used in migration research [45]. The second type of data is scientific publication data. This is a novel means of studying migration, concentrating on migrant scientists, and is based on the career trajectory as viewed from the publication record. Scientific migration flows give indications on the migration trends of highly-skilled migrants, providing insights into the brain drain/gain/circulation phenomena [45]. The third type of data we employ is Twitter data. Mobility analysis on Twitter is based on geo-location of user tweets and can provide real-time information on mobility trends [11, 26]. We use yearly flows on Twitter to observe migration trends. The fourth dataset consists of air passenger data, providing monthly and yearly passenger flows between countries. In these data, we can observe long-term, short-term, and seasonal migration [21]. Monthly aggregation reduces the effect of short traveling trips, while yearly aggregation reduces the effects of seasonal and short-term migration. As the fifth and final dataset, we analyze the highly-skilled migration from the Crunchbase platform.⁴

For each dataset, we calculate two mobility indicators. The first one, the *Flow Log Ratio*, shows the ratio between the number of people moving into the UK and the number of people moving out of the UK and gives a point-wise indication of the balance between the two flows. The second one, the *Cumulative Flow Log Ratio*, indicates the balance between the total number of incoming and outgoing people in the period analyzed.

Further, to better explore the components of our mobility indicator, we perform the time-series decomposition of the FLR values calculated from air passengers data. We investigate the pattern of the main trend over time, linking it to attention to Brexit using linear regression and Google Trends data.

Background and literature review

The study of migration has historically relied on data and statistics from national and super-national institutions and international organizations. In this context, so-called “traditional” gaps in international migration data have persisted over time [1]. These have been recently identified and classified into five categories, including (i) definitions and measurements, (ii) drivers or reasons behind migration, (iii)

⁴ Crunchbase, <https://www.crunchbase.com/>.

geographic coverage of the data, (iv) gaps in demographic characteristics and (v) the time lag in the availability of data [1]. The spatial coverage of data typically varies based on countries' development, with some leading countries collecting high-quality data and developing countries being less documented, if at all, undocumented. Similar to coverage, the timeliness of data varies across statistics and countries. The time lag in the data can vary from 2 to 20 years, significantly affecting the timeliness and actuality of the research.

Considerable efforts have been made to conceptualize, identify and bridge the gaps within and between traditional data and sources [1, 7, 45]. Efforts have led to improvements to some extent but to understand international mobility and migration patterns, develop scenarios, and design effective policies more accurate, reliable, and timely data are needed [1]. Nowadays, innovative data sources offer new opportunities to investigate mobility and migratory phenomena from new perspectives. Non-traditional data (also called *innovative data*) comes from individuals' digital footprints and sensor-enabled objects and includes social media and internet services data [20, 46, 52, 53], Call Detail Records (CDR) [42], air traffic [23], purchase transactions and money transfers [3]. We are at the moment when traditional and innovative data are jointly exploited in order to fully understand the complex and multifaceted nature of the migratory phenomenon and of mobility. Innovative data is most extensively used to fill in the gaps in traditional statistics and add new aspects typically overlooked/not captured by traditional data, such as social connections. The literature has shown that innovative data has great potential that sometimes overcomes measurement errors in survey data [34]. Specifically, innovative data typically show greater spatial and temporal coverage and granularity and are available—almost—in real time [34, 45]. However, similarly to the traditional context, innovative resources vary in terms of coverage and type of information, and pose some critical issues, including selection bias, and privacy and ethics [34, 42, 45].

In this study, we integrate traditional statistics (from Eurostat) and innovative data, including Twitter, scholarly data, Crunchbase data, and air traffic data, to investigate whether and how Brexit affects mobility and migration. We introduce two indicators based on log flow ratios, that allow us to eliminate short-term trips and observe the general trend in longer-term mobility. The use of flow ratios instead of net migration has the advantage that trends can be compared across different countries since the effect of the size of the country is removed when computing ratios. We study both monthly and yearly indicators, concentrating on a specific case study: Brexit and its effects on mobility to and from the UK. Also, this study employs time-series decomposition to extract the main trend from the flow ratios allowing us to observe the evolution of mobility between countries and geographical areas, capturing how it is affected by changes, including socio-political ones, such as Brexit.

Investigating Brexit

Given its political, economic, and social importance, multiple fields of study have investigated the Brexit phenomenon and its possible consequences. A review of scientific

publications focusing on the dynamics between Brexit and migration was published by [5]. According to this work, over three-quarters of the articles are based on original empirical data. Qualitative research methods, mainly interviews, prevail [5, 30]. Only a small number of articles propose quantitative research methods, and these are primarily based on surveys. Even fewer are the mixed methods (qualitative and quantitative). Almost all of the papers reviewed concern the UK exclusively—even with a further national or sub-national focus—and sometimes in combination with other EU or non-EU countries. Qualitative studies mainly discuss the reactions, perceptions, and experiences of EU nationals migrants living in the UK to the leave outcome [5, 24, 30, 49]. Moreover, studies discuss possible scenarios of UK migration policies after exiting the EU and the impact on the British and European economies [24, 39, 44]. Most of the studies focusing on EU migrants in the UK consider specific groups, such as citizens of Central and Eastern European countries, with particular attention to Polish and Bulgarian [5, 24, 30, 49].

Many researchers proposed both qualitative and quantitative studies assessing the “costs-of-non-Europe” [9, 32]. In terms of output or income, most studies highlight considerable economic costs from Brexit, with decreases ranging (in most cases) between -1% and -10% [8, 9, 38]. The analysis proposed in [48] focused on the variations in migration flows pre and post-Brexit 2016 referendum. Statistics show that EU immigration fell substantially after the Brexit referendum, while in the period before the referendum (2013–15), the UK saw an increase in the migration of EU citizens. The decline in migration to the EU in the post-referendum period mainly affected migrants from the new EU Member States, such as Poland.

This paper investigates whether and how Brexit has affected mobility trends between the UK and Europe in the last decades. As opposed to existing works, the study considers various types of mobility and demographic groups from official and innovative data i.e., Eurostat database, scientific publication, Twitter data, air passenger data, and Crunchbase data. Furthermore, we study the link between the attention to Brexit and migration topics from Google Search, and the mobility flows observed.

Mobility indicators

In this work, we are interested in studying the balance between incoming and outgoing mobility in the UK, and we employ two indicators: the *Flow Log Ratio FLR* and the *Cumulative Flow Log Ratio CFLR*.

The *Flow Log Ratio* is defined as the logarithm of the ratio between the number of incoming individuals (e.g., entering the UK) divided by the number of outgoing individuals (e.g., leaving the UK). Specifically, for a certain country or set of countries of destination C , and over a specified period of time, t , we consider the incoming flow $F_{C \rightarrow UK}(t)$ and the outgoing flow $F_{UK \rightarrow C}(t)$. The Flow Log Ratio $FLR(t)$ is then defined as

$$FLR(t) = \log_2 \frac{F_{C \rightarrow UK}(t)}{F_{UK \rightarrow C}(t)} \quad (1)$$

Table 1 Mobility datasets employed in this study

Dataset	Coverage	Flow time resolution	Type of mobility
Eurostat	2007–2019	Yearly	Long-term migration
Scientific publications	2000–2018	Yearly	Highly-skilled migration
Twitter data	2016–2022	Monthly and yearly	Long-term migration
Crunchbase data	2009–2022	Yearly	Highly-skilled migration
Air Passenger data	2011–2021	Monthly and yearly	Long-term, short-term, seasonal migration

Please note that the Brexit referendum was held on the 23rd of June 2016

A *FLR* below 0 indicates that more individuals moved out of the UK compared to those who moved in, while a value above 0 shows that the UK is an attractive country with more people coming in. A *FLR* of 1 means the incoming flows are twice as large as outgoing flows, while a *FLR* of -1 means the outgoing flow are twice less. The *FLR* is an indicator that allows us to study the trends point by point in time, and observe point-wise changes in trends. We are, however, also interested in a more general analysis over the entire period of interest. Therefore, we employ a second indicator, the *Cumulative Flow Log Ratio*, *CFLR*. This is defined as the logarithm of the ratio between the cumulative incoming flows and cumulative outgoing flows up to the current time window t :

$$CFLR(t) = \log_2 \frac{\sum_{t_i \leq t} F_{C \rightarrow UK}(t_i)}{\sum_{t_i \leq t} F_{UK \rightarrow C}(t_i)} \quad (2)$$

This second indicator allows us to understand long-term trends: whether the number of individuals moving into the UK is overall larger than those moving away, and what is the balance over recent years. Short point-wise changes in trend may not be significant at this level. However, a long-term change becomes very clear.

We study the two indicators at the European Union level and for European Union sub-regions, aggregating the flows among the countries in those regions. We use the division of EU member states into regions proposed by the EuroVoc vocabulary⁵: Northern (Finland, Denmark, Sweden, Estonia, Latvia, Lithuania), Southern (Greece, Italy, Malta, Portugal, Cyprus, Spain), Western (France, Germany, Ireland, Luxembourg, Netherlands, Austria, Belgium), Central and Eastern (Hungary, Poland, Romania, Bulgaria, Croatia, Slovakia, Czechia, Slovenia).

The time windows that we employ are yearly for all datasets, adding monthly analyses for Twitter and air passenger data which are the two cases where we have micro-data available.

⁵ <https://eur-lex.europa.eu/browse/eurovoc.html?params=72,7206>.

Datasets

This study exploits the potential of both traditional and innovative data from several mobility datasets covering different time periods and with different time resolutions to obtain a global view of mobility trends and to understand whether and how Brexit has affected mobility from different viewpoints. Table 1 summarises the mobility data characteristics.

Eurostat data

Eurostat,⁶ the statistical office of the European Union, provides a wide range of statistics on migration gathered through national statistical offices and EU-survey, e.g., the European Union Labor Force Survey. Among the datasets released by Eurostat,⁷ we focus on data derived from Demography and migration,⁸ which includes immigration and emigration flows, and indicators concerning migration stocks, e.g., foreign-born, citizenships, and marriages. Further, regarding migrants, data concerns (a) people born in a country different from the country of residence and (b) people with citizenship of a country other than the one of residence.

Eurostat data about bilateral migration flows constitute official data from national statistics, hence it is usually gathered either from national censuses and surveys or from population registries. When census data is not available, some values are provided by Eurostat as “estimated” or “provisional”, which means computed from mathematical and computational models or from other available statistics on total births and deaths [16]. We consider these data in our work.

Since our interest lies in inflows and outflows related to the UK, we extract from Eurostat:

- Flows by residence from EU27 to the UK (immigration) and return (emigration) [13, 15].
- Flows of UK citizens moving to the UK (immigration) and leaving the UK (emigration), from data based on citizenship [12, 14].
- Flows of EU27 citizens moving to the UK (immigration) and leaving the UK (emigration), from data based on citizenship [12, 14].

The temporal coverage spans from 2007 (or from 2013 in case of residence-based flows, when previous years were not available), up to 2019, year after which the UK flows were not updated, probably due to Brexit itself .

⁶ Eurostat: <https://ec.europa.eu/eurostat>.

⁷ Eurostat provides data on three main areas: Demography and migration, Asylum and managed migration (Eurostat, 2020: <https://ec.europa.eu/eurostat/web/asylum-and-managed-migration/background>), and Migrant integration. (Eurostat, 2020: <https://ec.europa.eu/eurostat/web/migrant-integration/background>).

⁸ Eurostat, 2020: Your key to European Statistics: <https://ec.europa.eu/eurostat/web/population-demography-migration-projections/background>.

Scientific publication data

Data relating to academic migration flow derives from the Enhanced Microsoft Academic Knowledge Graph (EMAKG) dataset [37]. This is an extension of the Microsoft Academic Knowledge Graph (MAKG) [18], a high-coverage dataset of publications and related entities, including authors (i.e., academics) and affiliations. Among other data extensions and annotations, EMAKG includes yearly worldwide flows of academics. For the scope of this study, we consider academics' country-to-country yearly flows from 2000 to 2018⁹

Twitter data

We provide real-time mobility trends using Twitter data. Using the Twitter API,¹⁰ we began by extracting all tweets sent from the UK in January 2015. In this period, we identified a total of 287,502 users. For all these users, we then downloaded the time and geo-location of all their tweets from January 2016 till November 2021. In this work, we focus only on the tweets with geo-location, as this is an important piece of information in order to observe the movements of the users. Hence, tweets without geo-location are ignored. Additionally, as the focus of this study is on the UK and European countries, we filtered out the tweets accordingly. We acknowledge that our choice of ignoring tweets without geo-location information may introduce bias. Selection bias is a known issue with Twitter data, and filtering by geo-location can enhance it Kim et al. [25]. However, Twitter has been demonstrated as an important data source in mobility studies, especially for early warning or to observe events in a timely manner Martín et al. [31], Ahmouda et al. [2]. Considering tweets with geo-location is the only means of assigning a location to a tweet with maximum accuracy. While other data such as Places could be used, these do not necessarily indicate location but also the topic discussed, and automatically distinguishing the two cases is not straightforward.¹¹ There are also systems that attempt to predict the location of the tweet, but again these do not have 100% accuracy [28, 40]. All these considerations have led us to choose geo-located tweets only. This reduces the amount of data, however our indicators do not look at in and out volumes but at ratios, which may alleviate the problem. We expect that the trends in the ratios we observe on the Twitter population that employs geo-location are indicative of the actual trend in the general population. This is also demonstrated by the fact that the trends match across datasets, therefore supporting the use of Twitter data for this application.

With these data, we then calculate the monthly and yearly UK mobility indicators. To be more specific, for each user, we first consider their most visited country as the usual country of residence for each month. Some users who tweet less

⁹ We omit the years 2019 and 2020 because of data incompleteness [19, 37].

¹⁰ <https://developer.twitter.com/en/docs/twitter-api>.

¹¹ "Tweets associated with Places are not necessarily issued from that location but could also potentially be about that location." from <https://developer.twitter.com/>.

or do not have the geo-location active will have missing country of residence for some months. We filter out the users that do not have the labels of usual country of residence for more than 50 months. This duration was selected to encompass a sufficiently extensive period to observe patterns, changes, and trends related to long-term migration behaviours on the platform. Through this choice of 50 months, we mainly observe cases where few months are missing in between months, effectively excluding scenarios characterised by significant lapses in monthly information. For other users, if we obtain a month without a residence location, we consider the residence to be that of the last month (i.e. we assume the user has not changed location). We acknowledge that there could be instances where users may change their location or move during the observed time frame. However, by employing this methodology, we aim to capture the overall trends and behaviours related to long-term migration for each user throughout the entire duration. This approach allows for a nuanced exploration of migration patterns over an extended time frame. With this procedure, we were able to identify movements of 60,366 users from/to the UK and from/to European countries from 2016 to 2021. We aggregate data of these users to compute both the monthly and yearly UK mobility indicators as set out in Eqs. 1 and 2 for movements between the UK and the four regions of EU member states.

Crunchbase data

Data from Crunchbase¹² regarding user and company information was obtained through the Crunchbase Academic Research Access Program.¹³ We employed these data to extract highly-skilled migration flows by aggregating users based on their estimated nationality and the headquarters of their workplace as the residence. Nationality/citizenship is assigned to a user by using the place of education as a proxy. For each user we consider the country where he/she has completed the studies as the country of nationality. Regarding residence, for each year a user is assigned a residence based on the country they worked in in January. Flows are computed from 2009 to 2022 by counting the users who changed jobs in a specific year, and the new and old jobs are in different countries. In this way, a user who changes countries more times in a year is counted in all the corresponding flows, which is desirable when studying highly-skilled migration.

Air passenger data

We used monthly air passenger traffic volumes from each country of origin and to the true final country of destination provided by the Sabre Corporation [41]. This dataset describes the total number of passengers flying between two countries, regardless of whether the flights are direct or indirect. The dataset covers worldwide

¹² Crunchbase: www.crunchbase.com.

¹³ <https://about.crunchbase.com/partners/academic-research-access/>.

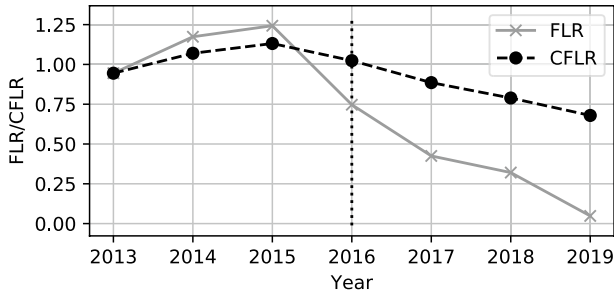


Fig. 1 *FLR* and *CFLR* calculated from Eurostat flows, taking into account all changes in residence regardless of nationality. The vertical line shows the year of the Brexit referendum

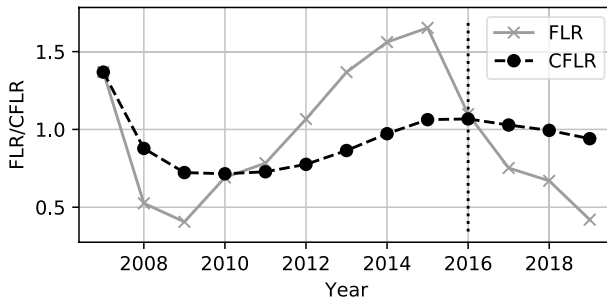


Fig. 2 *FLR* and *CFLR* for EU27 citizens calculated from Eurostat flows. The vertical line shows the year of the Brexit referendum

air traffic between February 2011 to October 2021. The air traffic indicators computed are available at <https://doi.org/10.7910/DVN/AE1PKC>.

Mobility trends

For each dataset in Table 1 we extracted UK incoming and outgoing flows and computed the values of the two indicators. In this section, we discuss the values obtained for each dataset.

Migration trends from Eurostat data

For Eurostat data, we studied migration flows at the EU level, i.e. the number of persons entering the UK from EU27 countries and leaving the UK towards EU27. Unfortunately, the data on individual countries had missing values, therefore we could not compute the indicators at the regional level as we did with the other datasets.

Figure 1 displays the two indicators from 2013 to 2019. We observe both for *FLR* and *CFLR* a strong inversion in trends starting from 2016. If up to 2015

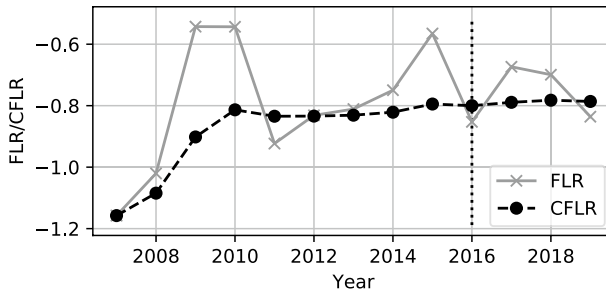


Fig. 3 *FLR* and *CFLR* for UK citizens calculated from Eurostat flows. The vertical line shows the year of the Brexit referendum

the UK was a strong attractor of migration in Europe, with immigration more than twice as large as emigration, after 2016 the ratio started to drop, reaching almost 0 in 2019, therefore incoming and outgoing flows become comparable. The cumulative indicator shows that still, until 2019, the overall balance over the year 2013–2019 was inclined towards larger immigration (value slightly under 0.7). However, if the trend continues, it is possible for the emigration of UK residents to take over in the next years.

In order to get a better understanding of migration trends, we also studied the indicators for two sub-populations of migrants. First, we looked at EU27 citizens moving in and out of the UK, and computed the *FLR* and *CFLR*, as shown in Fig. 2. In this case, we were able to obtain data for the period 2007 to 2019. The difference from Fig. 1 is that UK citizens and non-EU nationals are not included here, hence we can observe the effect on EU citizens only; on the other hand, the destination of these citizens is not necessarily towards the rest of the EU, but it can be towards any world country. The general pattern observed is the same: from 2009 up to 2015 the attractiveness of the UK for EU27 citizens was steadily growing, with more than 3 times more immigrants than emigrants in 2015. However, from 2016 the *FLR* dropped significantly, being below 0.5 in 2019. The cumulative ratio is also decreasing, however still almost twice as many EU27 citizens entered the UK compared to those that exited in the period 2008–2019 (*CFLR* slightly below 1 in 2019).

A second sub-population studied is that of UK nationals, for which Fig. 3 shows the two indicators. We note that the *FLR* values are below 0, meaning that the number of *UK emigrants* is generally larger than that of UK nationals returning to the UK. The values we observe are fluctuating around -0.7 , without a definite change in trend after 2016. However, by looking at the cumulative trend, we see a more clear picture. While in the early years (2007–2008) the number of emigrants was much larger than that of returners (more than twice—*CFLR* under -1), in time the difference has been decreasing, with a local peak in 2010. Between 2010 and 2014 the trend was stable. However, in recent years (2015–2019) the ratio started again to increase, albeit slowly, indicating that more and more UK nationals are returning to the UK, and less are emigrating. The balance, however, is still in favour of emigrants.

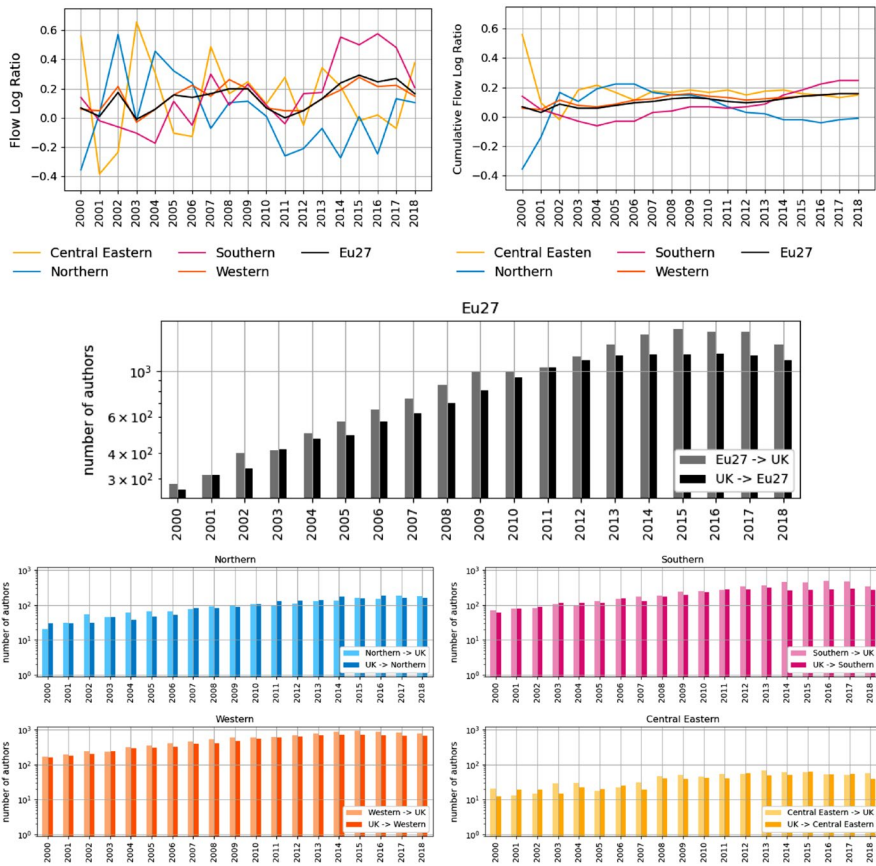


Fig. 4 Top: *FLR* and *CFLR* calculated from scientific flows between UK and EU27 and EU27 regions. Bar plots: Incoming and outgoing scientific flows for UK versus EU27, Northern, Southern, Western and Central Eastern European countries

Scientific migration

Another analysis concentrates on migration flows for scientists extracted from publication data. This allows us to study the mobility of highly-skilled individuals and can give indications about the brain circulation phenomenon. Figure 4 (top subplots) shows the value of the two indicators obtained from publication data. The EU-level *FLR* values show values generally above zero indicating that the UK is an attractor also for scientific migration, with more scientists moving from EU27 countries to the UK than the opposite. An increase in this effect is visible in the periods 2005–2009 and 2012–2015. From 2016 we observe a plateau and a decrease in *FLR* in 2018. The *CFLR* confirms that the UK has been steadily gaining scientist migrants over the years, with values above zero and increasing up to 2017. In 2018 it seems that the values have stabilised.

Since we are observing a very specific population, the number of persons included in the study is not very high, especially for some EU sub-regions. To avoid misinterpretation of *FLR* and *CFLR* values, we also display the volume of the incoming and outgoing flows in Fig. 4 (bar plots). At EU27 level, these show how scientific migration to and from the UK has been steadily growing between 2000 and 2015 from the volume point of view. From 2016 we observe a plateau and then an inversion in trend in 2018.

We also study regional data:

- Western Europe follows the same trend as EU27, both in terms of ratios and volumes, being in fact responsible for most of the flow volume.
- Southern Europe also follows the general trend, however here changes are sharper. We observe an important increase in the years just before 2016, especially in terms of emigration towards the UK, with a strong inversion in trend in 2017 and 2018, visible both in terms of *FLR* and volumes. The *CFLR* shows a stable trend for 2017 and 2018.
- Central and Eastern countries show a lot of variability in the *FLR*, with most values greatly above zero and an apparent decrease from 2015 to 2017, and again a large value in 2018. However, when we combine this information with the flow volumes, we observe that there are rather limited, so the fluctuations we see in terms of *FLR* could be simply noise. We note an important increase in volumes (both in and out) from 2008 on (after Romania and Bulgaria have joined the EU in 2007) and a decrease or stabilisation of volumes from 2016 on. The *CFLR* shows less fluctuation and confirms that the UK has been a steady attractor for scientists from central and Eastern Europe, with a slight inversion in trend after 2015.
- Northern Europe shows a different trend compared to other regions. While in the early 2000 s *FLR* and *CFLR* values were above zero, showing that the UK was an attractor of scientific migration, these have been slowly decreasing after 2006. The *FLR* stayed below zero almost continuously between 2011 and 2016, while the *CFLR* fell below zero in 2014, showing that the number of individuals who moved out of the UK since 2000 has overcome those who moved into the UK. A slight inversion in trends appears to happen in 2017 and 2018 for both *FLR* and *CFLR*. The volumes follow an increasing trend up to 2015 and enter a plateau afterwards.

All in all, there are clear indications that around 2016 there was a change in scientific migration between the EU and the UK. While before 2016 volumes of migration in both directions were steadily growing, afterwards they started to stabilise or decrease. The *FLR* indicator shows that up to 2015 the UK was a strong attractor of highly-skilled migrants from Europe, especially Southern but also Western Europe, and that afterwards the trend started to stabilise and decrease. Northern Europe instead seems to display the opposite trend. The *CFLR* confirms these findings also on the longer term.

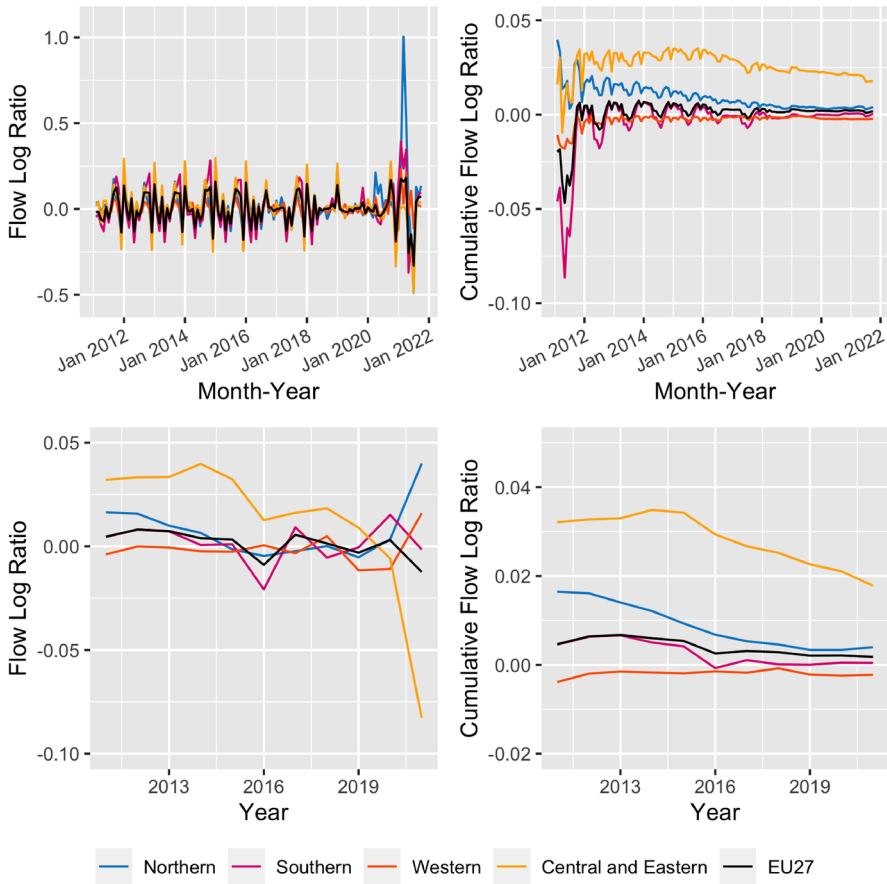


Fig. 5 Monthly (top figures) and Yearly (bottom figures) *FLR* and *CFLR* from Air Passenger data

Mobility from air passenger data

Using Air Passenger data [41] we estimated the monthly and yearly *FLR* and *CFLR*, displayed in Fig. 5. For the yearly indicators, we estimated the passenger volumes from the beginning of February to the end of January of the next year, to avoid the seasonal effects of the Christmas and new year holidays, since our data starts in February 2011.

It is important to note that, compared to indicators from Eurostat and scientific data, which allowed us to measure migration directly, *FLR* and *CFLR* values extracted from Air Passenger data are much closer to 0. This is due to the fact that Air Passenger data measures overall mobility, and tourism volumes are much higher than migration volumes, which explains the low ratios, especially for yearly and cumulative indicators. The average UK monthly flow in these data is about 4.5 million passengers, both incoming and outgoing, with an average absolute net volume

of about 180,000 passengers. Yearly figures are even more pronounced: more than 50 million passengers, on average, per year, while absolute net volumes remain around 180,000 (possibly migrants). Even though much smaller than total volumes, these net volumes are larger than the other non traditional datasets in this paper, therefore we believe this dataset is very representative and central to our analysis. It is important to underline that tourism interferes only with the range of our indicators, while we are interested in the changes in time, which we observe to be similar across datasets.

The yearly indicator shows patterns similar to what we observed in previous data. The EU27 *FLR* values are always above 0 in the years preceding 2016 and then they start to oscillate, reaching values below 0 in 2016, 2019, and 2021. Similar to Eurostat data, we observe a slight increase in early years, and a sharp decrease in 2016. While Eurostat data continues to indicate a decrease from 2017 on, here we see an oscillating pattern. At the regional level, the largest change in trend is observed with Central and Eastern Europe, where *FLR* dropped in 2016, and again after 2019. Southern Europe also displays a strong decrease in 2016 and then oscillations afterwards. Similar to previous data types, Northern Europe appears to have started a decrease in *FLR* long before 2016, and shows a high peak in 2021. When looking at *CFLR* values, the overall trend is downward for all countries after 2016, indicating that slowly more people are travelling from the UK to Eu27 countries than vice versa. The EU27 trend is similar to Eurostat data. All these values indicate that Brexit has probably started to have an effect in 2016.

A very important and unique feature of the air passenger data is the fact that it allows us to study monthly patterns as well, which is not the case for most other datasets, including official data. We note a strong seasonality in the monthly data, with a typical high peak in *FLR* in January, April and September, with more individuals entering than exiting the UK, and low peaks in March, May, July and December, with more individuals exiting than entering the UK. This pattern could be explained partly by tourism, for instance July could be the summer holiday time when many UK residents go abroad to return in August or September. It has been previously shown that summer holidays correspond to higher airline traffic [27]. Similarly, many residents (migrants or not) could be travelling abroad in December to return in January. However, seasonal migration can also be responsible for this pattern [21]. For most of the last two decades, the EU was the primary source of migrant workers in the UK, particularly in the field of agriculture¹⁴ and horticulture. For example, UK in a Changing Europe¹⁵ estimates that in fruit picking seasonal migrant workers (in the UK from the beginning of the year until November), constitute up to 98% of workers.¹⁶ The seasonal pattern is also visible in the *CFLR*, with periodical oscillations at the EU27 level. For Central and Eastern Europe, it is clear that over time the UK is an important attractor, with values always above 1, and we see a decrease in

¹⁴ The Migration Exploratory: <https://migrationobservatory.ox.ac.uk/resources/briefings/work-visas-and-migrant-workers-in-the-uk>.

¹⁵ UK in a Changing Europe: <https://ukandeu.ac.uk/>.

¹⁶ <https://ukandeu.ac.uk/seasonal-workers-visas>.

CFLR from July 2016 on. Western Europe appears to show much weaker seasonal patterns, indicating that seasonal travelling and migration could be stronger in the other regions. The decreasing trend for Northern Europe, starting already in 2011, is confirmed by these data as well. Southern Europe, on the other hand, seems to be quite balanced, oscillating around 1 during the year.

An important aspect that these data allow us to observe is that, while before 2016 the *FLR* and *CFLR* seasonal patterns are very strong and stable, after 2016 they disrupted. In particular, for *FLR* the December 2016/January 2017 peaks are missing, while for 2018 and 2019 all peaks are missing except for January. The *CFLR* curve shows some disruptions in 2016 and 2017, and a strong flattening after 2018. This disruption in periodic patterns is very interesting and could be explained in part by the lower tendency to travel for UK migrant residents, for fears of not being able to reenter the UK. Also, seasonal migration could have been disrupted by Brexit events and media coverage.

Another important aspect concerns the period after January 2020, where we also observe important changes in trends and sometimes very high oscillations. These disruptions are a sum of Brexit and Covid19 effects and are difficult to assign to one or the other. We note flattening periods and stronger peaks in this period, especially for Northern and Southern Europe, probably due to bursts in mobility after lockdown periods. This cumulative effect makes it difficult to extract meaningful information on Brexit during this period, however the disruption is clearly visible in the years before (2016–2019), therefore these data still provide an important contribution to our study. The *CFLR* stabilises after 2020 indicating that in this period seasonal mobility is not strong, again probably a cumulative effect of Covid19 and Brexit. The decreasing pattern continues for Central and Eastern Europe, indicating that many individuals leave the UK towards these countries and fewer enter the UK.

Mobility through Twitter data

Twitter is another type of data that allows us to explore general mobility patterns including short-term travelling, seasonal migration, long term migration, with the advantage of a high space and time resolution, therefore it has been widely used for mobility studies [25, 31]. Unlike Air Passengers data, Twitter data is more easily accessible, through the Twitter API, while Air Passenger Data requires ad hoc agreements with the data owners. Twitter includes a much smaller population, referring only to Twitter users that have their geo-location enabled. However, the patterns can still be of interest as they show trends for this sub-population. Therefore we also include this data type in this study, to analyse the patterns and also evaluate its suitability to this type of study, especially useful as this data has recently stopped being free.

Figure 6 shows *FLR* and *CFLR* values obtained by observing the movements of users who were in the UK in January 2015, so before Brexit. At the EU27 level, yearly values of both indicators are generally below 0, indicating that more users have left the UK compared to those coming back. In 2020 and 2021 we see fluctuations in these values, especially for *FLR*, but these could be due to Covid-19 effects.

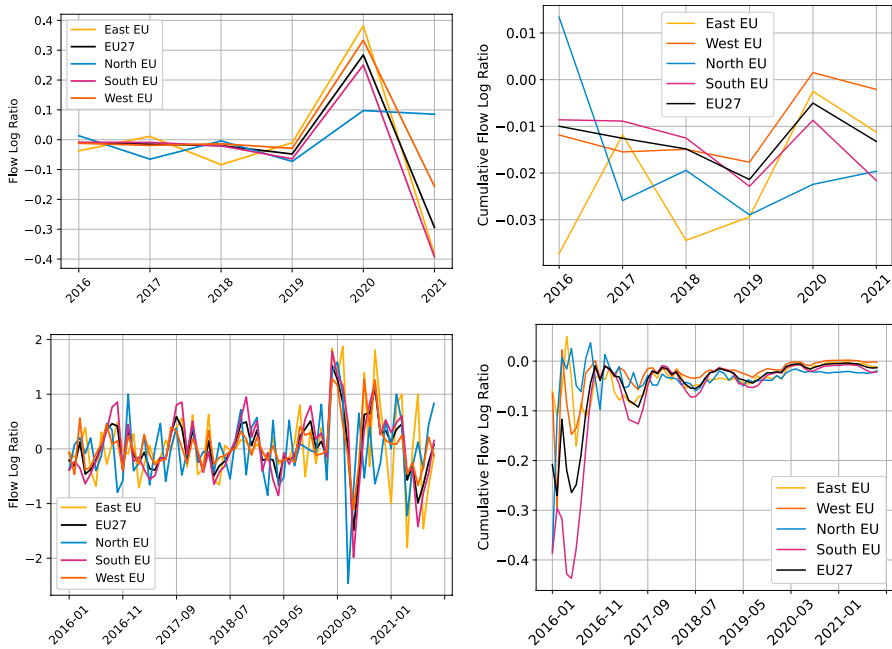


Fig. 6 Yearly (top panel) and monthly (bottom panel) *FLR* and *CFLR* between the UK and the sub-regions of EU on Twitter

Analysing individual regions, we observe that Northern Europe show the lowest *FLR* and *CFLR* values, with a clear trend for users to leave the UK in this period, in agreement with what we observed for other data types. Western and Southern Europe are well align with the EU27 trend, while Central and Easter Europe show high fluctuations, most probably due to the low number of users observed in this area.

Regarding monthly values, the data provide a much more useful picture. Twitter *FLR* also shows a regular pattern over the years, however not all peaks observed in Air Passenger data are visible here as well. For instance, we have a peak in users entering the UK in September/October, which matches Air Passengers data. However, the January peaks are missing here, in years where they are missing in Air Passenger Data (2017, 2019, probably due to Brexit), but also in 2016 and 2018. This could indicate that the individuals who travel in December and January are generally not Twitter users, belonging to demographic groups that are less represented. We see low *FLR* values during the summer, indicating that more users are leaving compared to those entering the UK. The autumn peak could be due to students coming from the EU to the UK to study, or tourists coming back from holidays (especially visible for Southern Europe) while the summer low values could be due to vacation travelling. We also observe strong disruption in patterns in 2020 and 2021, with a high peak in travelling from Northern Europe to the UK in February 2021, matching a similar peak in Air traffic data. Importantly, we see good agreement of monthly patterns over the years 2016–2020 (from Brexit referendum but pre-Covid19), as



Fig. 7 Top: *FLR* and *CFLR* calculated from Crunchbase flows between UK and EU27 and EU regions. Bar plots: Incoming and outgoing flows from Crunchbase for the UK versus EU27 and EU regions

opposed to monthly patterns in Air Passenger Data that are quite disrupted in the same period. This suggests that while the general population changed mobility due to Brexit, this is less visible in Twitter users, who are likely less conditioned by visas or labour instability.

All in all, although Twitter data is more restricted and subject to selection bias, we do observe some patterns similar to those of other data types. In particular, we observe different trends for different EU regions and periodic mobility patterns. This is a further support for the usefulness of these types of data that could provide quick feedback and early warnings on the effect of policies and other socio-economic changes, even if overall estimates are subject to the Twitter selection bias.

Highly-skilled mobility from Crunchbase data

In order to analyze possible Brexit effects from Crunchbase data, we focus on the United Kingdom by calculating mobility indicators for Crunchbase flows. Figure 7

shows EU27 and regional *FLR* and *CFLR* calculated for the UK, and also the total incoming and outgoing flows from/to European regions.

When studying the entire European Union (EU27), after an initial negative trend, the *FLR* settles around 0 from 2012 to 2015. From 2016 to 2019 it decreases and then rises again from 2019 onwards. Values are mostly negative, indicating that for the Crunchbase highly-skilled population, more people moved out rather than into the UK. The cumulative values (*CFLR*) confirm this trend. However, the UK increased its attractiveness from 2011 to 2016, with an upwards pattern. On the contrary, from 2016 to 2019, we observe a decreasing pattern, probably due to the first effects of Brexit. After 2019 the pattern seems to be increasing again. We must however note that flow volumes are very reduced, even when looking at EU27 level.

Focusing on single regions:

- West Europe shows a pattern similar to the European one.
- The *FLR* of Southern countries varies significantly yearly, passing from positive to negative values and vice versa. This noisy behaviour could be due to the low flow volumes. The *CFLR* for Southern Europe shows an increasing pattern from 2010 to 2016, meaning that the UK has become more attractive over time for highly-skilled individuals. While negative values prevailed in early years, when more highly skilled individuals were leaving the UK, positive values emerged from 2014, indicating that incoming flows prevailed. From 2017 onwards, the indicator starts decreasing again so that in 2022 it stands just below 0. This inversion in the pattern indicates that since the immediate post-referendum (and at least until 2020), the UK has become increasingly less attractive for highly skilled personnel from southern European regions.
- Northern European countries show generically negative *FLR* values just below 0. The *CFLR* confirms a stable negative pattern, indicating higher outgoing flows from the UK to Northern countries, compared to incoming flows. This means that from the point of view of highly-skilled workers, the UK is not attractive to Northern Europe.
- Central-Eastern Europe always shows a decidedly negative pattern, except for a positive peak in 2013. Consistently, the *CFLR* shows a decreasing trend from 2013 onwards, which stands at around -0.5 from 2017 onwards. This pattern could indicate, similar to Northern Europe, a low degree of attractiveness of the UK for highly skilled workers, which could also be due to poor international agreements. However, as shown in the bar plots of Fig. 7, we have extremely low flow volumes for this region.

All in all, the analysis indicates that after Brexit was approved (2017-2019), there was a general decreasing trend in *FLR* values for all European regions, showing that more highly-skilled individuals tended to leave the United Kingdom. However, the ratio has restarted to grow from 2020 on. We observe a stable volume decrease starting from 2017, for all regions except for Northern Europe. This is in line with what we observed from scientific data: Brexit appears to have partially encouraged incoming flows from

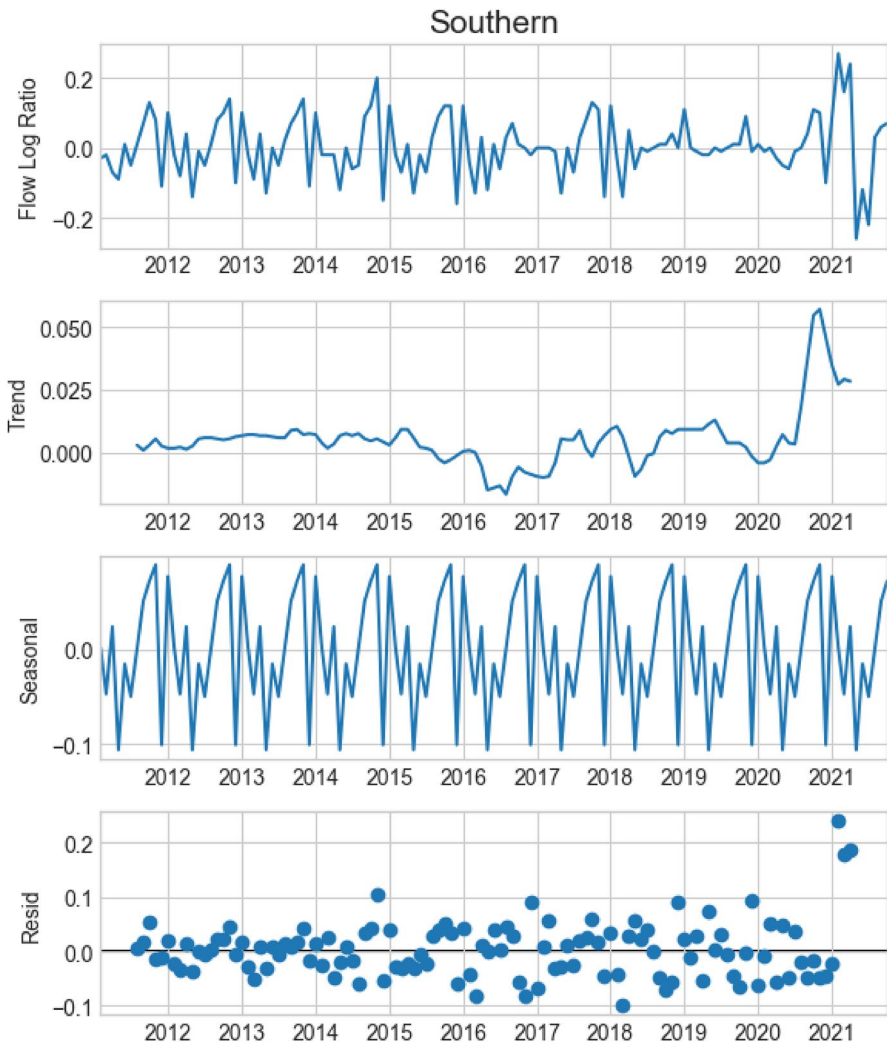


Fig. 8 Time series decomposition of *FLR* calculated from Air Passengers data for Southern European countries

North Europe and discouraged those from the other European regions. Even if overall conclusions are similar to other data types, we underline the fact that Crunchbase data is rather reduced in volumes, therefore data integration is always necessary to validate results.

Brexit attention and time-series decomposition of European Union air passenger data: analyzing main trends

In order to explore better the main trends and seasonal patterns, we perform time series decomposition on the *FLR* values calculated from Air Passenger data, similar to [22]. We dissect regional *FLR* values into three components: main trend, seasonal component and noise, by employing the Python library statsmodels [33]. We set a seasonal component for each month, for a total of 12 components. As an example, Fig. 8 shows the result obtained for Southern European countries: the top panel shows the original *FLR* time series, which reveals strong variations re-occurring quite regularly monthly every year until 2016 and in 2018. The main trend and seasonality are displayed in the second and third panel, respectively, while the residual component, i.e., noise, is shown in the bottom panel. The first and the last six months of the main trend are removed since time-series decomposition employs a moving window.

Analysis of main trends

Time-series decomposition allows us to study the main trend separately from the seasonal component. We compare this main trend with data from Google Trends, that measures the amount of searches on a topic of users in a specific country. The number of searches can be used as a measure of the attention of people in the country to a specific subject. It was widely investigated in several research fields including migration [6, 17]. To collect Google Trends data, we design a Python pipeline combining Pytrends, an unofficial API for Google Trends,¹⁷ and the bulk download method¹⁸ provided in [51]. According to our research question, data collection is restricted to European countries plus the United Kingdom from 2010 to 2021. For each country, we download time series indexes for “UK visa” (*UkVisa*), “UK residence permit” (*UkResPerm*), plus individual country names, as keyword terms, and *Brexit* as a topic. We aggregate (sum) countries’ Google trends based on European geographical areas, to obtain regional trends. For the terms related to country names, we perform two aggregations: one by grouping all the searches performed in the UK for the countries of a certain region, and one by grouping the search for the term ‘United Kingdom’ performed in all countries of a certain region.

Figure 9 shows the patterns of *FLR* main trends for the UK (after time series decomposition) versus the four European Union regions, and the “Brexit” Google Trends time series data for the UK. Again, we note that in some regions some peaks in Brexit attention coincide with main trend changes: however, there are no clear correlations between the two series.

Main trends extracted with the decomposition of time series are further studied with respect to users’ searches on Google Trends. For each European region, we

¹⁷ <https://pypi.org/project/pytrends/>.

¹⁸ <https://gist.github.com/pixyfox>.

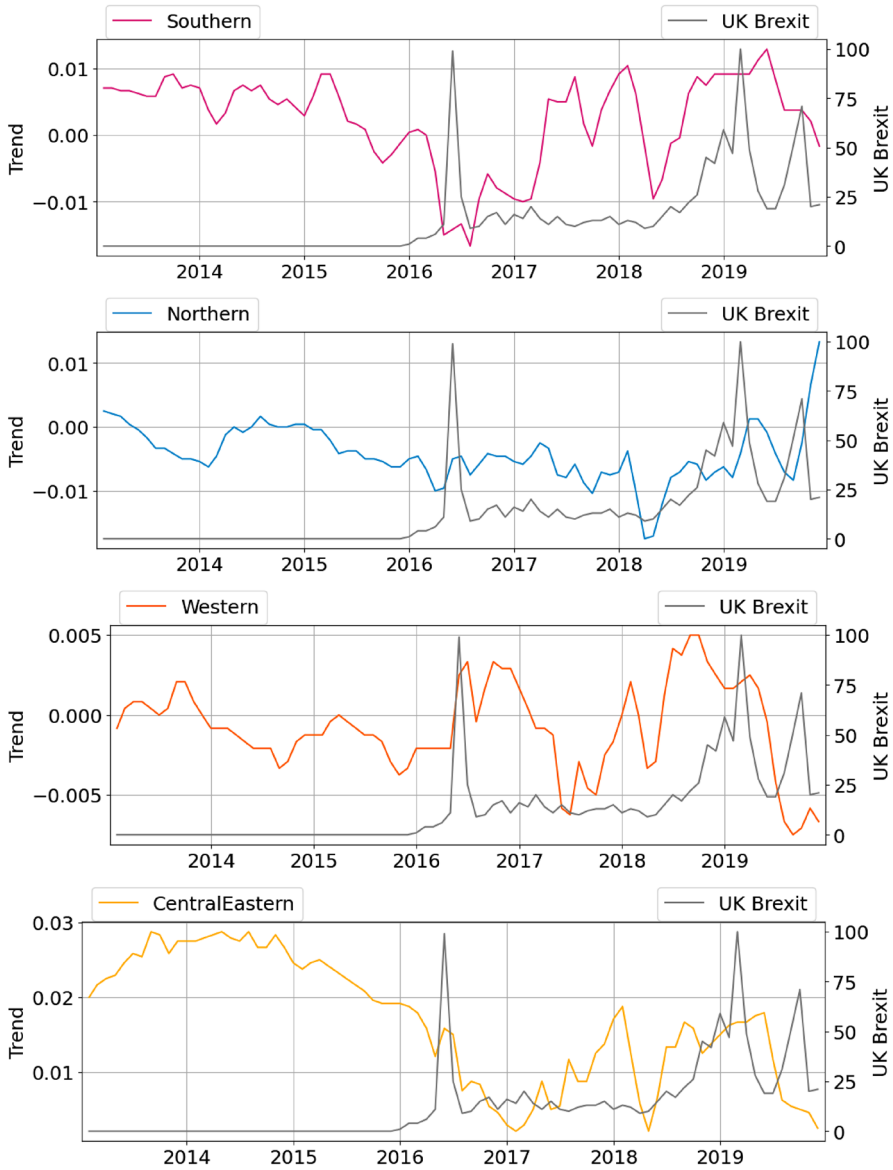


Fig. 9 Time-series comparing EU27 regional main trends and the "Brexit" Google Trends in UK

implement an Ordinary Least Squares (OLS) model with an automated backward elimination approach. The dependent variable considered is the monthly main trend of the *FLR*, for each region. Each model considers as the input variables (a) attention levels for different topics (UK Visa, UK Residence Permit and Brexit) derived from Google Trends data from both the European Region of interest and the UK, (b) attention to country terms pertinent to each region in the UK and attention

Table 2 OLS coefficients, estimated standard deviation, and R² and AIC for Northern and Southern regions when modelling trends from 2013-08-01 to 2018-12-01

Variable	Northern		Southern	
	Northern base	Northern final	Southern base	Southern final
UkVisa 6 m	0.0978 (0.221)		-0.2198 (0.311)	-0.3701** (0.132)
UkVisa 1 m	-0.0244 (0.082)		-0.0633 (0.119)	
UkVisa 12 m	-1.5825*** (0.457)	-1.3899*** (0.26)	-0.2495 (0.422)	
UkResPerm 6 m	-0.0534 (0.124)		0.1345 (0.167)	
UkResPerm 1 m	0.0483 (0.07)		0.0169 (0.078)	
UkResPerm 12 m	-0.0188 (0.154)		0.1559 (0.186)	0.231 (0.119)
trend y-2	0.7743** (0.233)	0.7511*** (0.174)	-0.0665 (0.307)	
trend y-1	-0.5036* (0.215)	-0.4939*** (0.138)	-0.4883 (0.347)	-0.2962 (0.174)
Brexit 6 m	0.2353 (1.576)		-1.4201 (1.638)	-0.4451* (0.196)
Brexit 1 m	-1.0246 (0.605)	-0.99** (0.319)	-1.6088* (0.703)	-1.5365** (0.506)
Brexit 12 m	1.8297 (1.765)	2.1201** (0.72)	3.9326 (2.036)	2.4653* (1.053)
United Kingdom 6 m	0.005 (0.171)		0.4911 (0.265)	0.6426** (0.196)
United Kingdom 1 m	-0.1389 (0.083)	-0.1285 (0.068)	-0.2032 (0.125)	-0.2424* (0.095)
United Kingdom 12 m	0.2504 (0.161)	0.2553* (0.099)	0.152 (0.398)	
UK UkVisa 6 m	-0.1079 (0.208)		-0.589 (0.467)	
UK UkVisa 1 m	0.1244 (0.092)	0.111 (0.058)	0.2005 (0.193)	
UK UkVisa 12 m	-0.1159 (0.661)		2.693** (0.792)	2.0969*** (0.439)
UK UkResPerm 6 m	-0.2343 (0.625)		-0.9561 (0.882)	-1.1643* (0.518)
UK UkResPerm 1 m	-0.4421* (0.201)	-0.5006*** (0.13)	-0.3377 (0.278)	-0.3211 (0.228)
UK UkResPerm 12 m	-1.1144 (0.793)	-1.2285*** (0.341)	-1.8428* (0.792)	-1.2929* (0.556)
UK Brexit 6 m	-0.583 (2.44)	-0.4064** (0.144)	1.4666 (2.614)	
UK Brexit 1 m	1.5416* (0.752)	1.5*** (0.386)	2.1516* (0.882)	2.073*** (0.613)
UK Brexit 12 m	-2.891 (2.879)	-3.1348** (1.036)	-6.3548 (3.874)	-3.7198* (1.689)
UK S-Countries 6 m			0.8444** (0.3)	0.5963*** (0.155)
UK S-Countries 1 m			-0.1838 (0.186)	-0.1436 (0.084)
UK S-Countries 12 m			-3.2354*** (0.889)	-3.302*** (0.709)
UK N-Countries 6 m	0.4999* (0.21)	0.4978*** (0.137)		
UK N-Countries 1 m	-0.1777* (0.084)	-0.1633** (0.06)		
UK N-Countries 12 m	0.8999* (0.421)	0.6147* (0.276)		
Backward AIC	77.9594	61.5	111.6961	99.5795
Backward R2	0.9154	0.9106	0.8578	0.8443

Coefficients with p-value ≤ 0.05 are flagged with one star (*), ≤ 0.01 with 2 stars (**), ≤ 0.001 with 3 stars (***)

Table 3 OLS coefficients, estimated standard deviation, and R^2 and AIC for Western and CentralEastern regions when modelling trends from 2013-08-01 to 2018-12-01

Variable	Western		CentralEastern	
	Western base	Western final	CentralEastern base	CentralEastern final
UkVisa 6 m	-0.6181 (0.48)	-0.5902* (0.275)	-0.1769 (0.097)	-0.1654 (0.091)
UkVisa 1 m	0.1056 (0.11)		0.0158 (0.05)	
UkVisa 12 m	-0.1213 (1.061)		0.2248 (0.127)	0.2979* (0.113)
UkResPerm 6 m	0.0257 (0.214)			
UkResPerm 1 m	-0.0303 (0.083)			
UkResPerm 12 m	-0.0363 (0.237)			
trend y-2	0.0562 (0.151)		-0.3017 (0.187)	-0.2498* (0.112)
trend y-1	-0.7284*** (0.164)	-0.7045*** (0.101)	-0.532* (0.229)	-0.5376*** (0.159)
Brexit 6 m	1.4733 (1.772)	1.3385 (0.812)	0.7327 (1.043)	0.9565 (0.748)
Brexit 1 m	0.3157 (0.671)	0.327*** (0.08)	-1.4562*** (0.356)	-1.3971*** (0.298)
Brexit 12 m	-4.0464 (2.546)	-3.6471*** (0.968)	2.6968* (1.031)	2.4761** (0.758)
United Kingdom 6 m	-0.5593** (0.201)	-0.5793*** (0.132)	0.1379 (0.189)	
United Kingdom 1 m	-0.0541 (0.11)		-0.0933 (0.06)	
United Kingdom 12 m	0.3628 (0.37)	0.299 (0.208)	-0.0076 (0.161)	
UK UkVisa 6 m	-0.9654* (0.384)	-0.8622*** (0.206)	-0.1301 (0.165)	
UK UkVisa 1 m	0.2707* (0.131)	0.2883** (0.087)	0.1928** (0.067)	0.1879*** (0.051)
UK UkVisa 12 m	0.5124 (0.739)	0.3391 (0.182)	0.2863 (0.285)	
UK UkResPerm 6 m	1.2915 (0.771)	1.3358* (0.509)	0.4843 (0.406)	0.6259* (0.245)
UK UkResPerm 1 m	-0.448 (0.245)	-0.4559* (0.185)	-0.2211 (0.126)	-0.1955 (0.12)
UK UkResPerm 12 m	0.0901 (0.677)		-1.1397 (0.58)	-1.2812*** (0.341)
UK Brexit 6 m	-2.1705 (2.611)	-1.8898 (1.182)	-2.2946 (1.579)	-2.4081* (1.074)
UK Brexit 1 m	0.0655 (0.81)		2.2086*** (0.446)	2.0888*** (0.371)
UK Brexit 12 m	5.618 (3.893)	5.0722** (1.636)	-3.1834 (1.793)	-2.9789* (1.406)
UK W-Countries 6 m	0.6599** (0.222)	0.5601*** (0.123)		
UK W-Countries 1 m	-0.0361 (0.106)			
UK W-Countries 12 m	-0.0795 (0.214)			
UK C-Countries 6 m			0.6496** (0.22)	0.5059*** (0.137)
UK C-Countries 1 m			-0.3567** (0.118)	-0.3684*** (0.099)

Table 3 (continued)

Variable	Western		CentralEastern	
	Western base	Western final	CentralEastern base	CentralEastern final
UK C-Countries 12 m			-0.8798 (0.553)	-0.6349* (0.286)
Backward AIC	103.791	84.463	29.9323	22.6419
Backward R2	0.874	0.8688	0.9557	0.9523

Coefficients with p-value ≤ 0.05 are flagged with one star (*), ≤ 0.01 with 2 stars (**), ≤ 0.001 with 3 stars (***)

to the United Kingdom term in the specific region, and (c) the main trend of the previous year and two years earlier—in the same month—($\text{trend}_{y-1, y-2}$) obtained by time-series decomposition. Regarding Google Trends, for each monthly observation, for each keyword/topic, we include as independent variable the number of searches in the previous month and the average of searches in the previous six and twelve months in the specific geographical area observed and in the UK. For some geographical regions the Google Trends time-series related to some topics are completely empty, in which case they are removed. Prior to model fitting, dependent and independent variables are standardised. The model takes all variables as input and, at each iteration, drops the least significant based on a fixed elimination criterion—here we employ the Akaike information criterion (AIC). When only the most statistically significant variables are left, the algorithm stops.

Tables 2 and 3 shows OLS coefficients, estimated standard deviation, and R^2 and AIC by geographical region when modelling trends from 2013-08-01 to 2018-12-01. We show results both for the model with all independent variables (base), and the final model after backwards elimination (final). Coefficients with p-values less (or equal) than 0.05 are flagged with one star (*); less (or equal) than 0.01 are flagged with 2 stars (**); less (or equal) than 0.001 are flagged with three stars (***). In brackets, we include the estimates of the standard deviation of the coefficient, a measurement of the amount of variation in the coefficient throughout its data points (*standard error*). Empty cells in the base models' columns', i.e., UkResPerM averages in the previous 1/6/12 months in Central Eastern, indicate the completely empty Google Trend series removed before computing the OLS model. Further, empty cells in the final models' columns', e.g., UkResPerM averages in the previous 1/6/12 months in Southern, Central Eastern and Western Europe, indicate that the algorithm deleted the features during the iterations and thus are not included in the final model.

In general the OLS models obtain a very good fit, with R^2 values above 0.84. This demonstrates that the independent variables are able to describe the dependent variable, i.e. there is a good relation between Google trends and previous values and the current monthly values we are modelling.

For a global view of the models' performance, the line plots in Fig. 10 shows the actual trends (black dotted line) and the trends predicted (coloured lines). For all European regions, the two series are very similar, as indicated also by the high

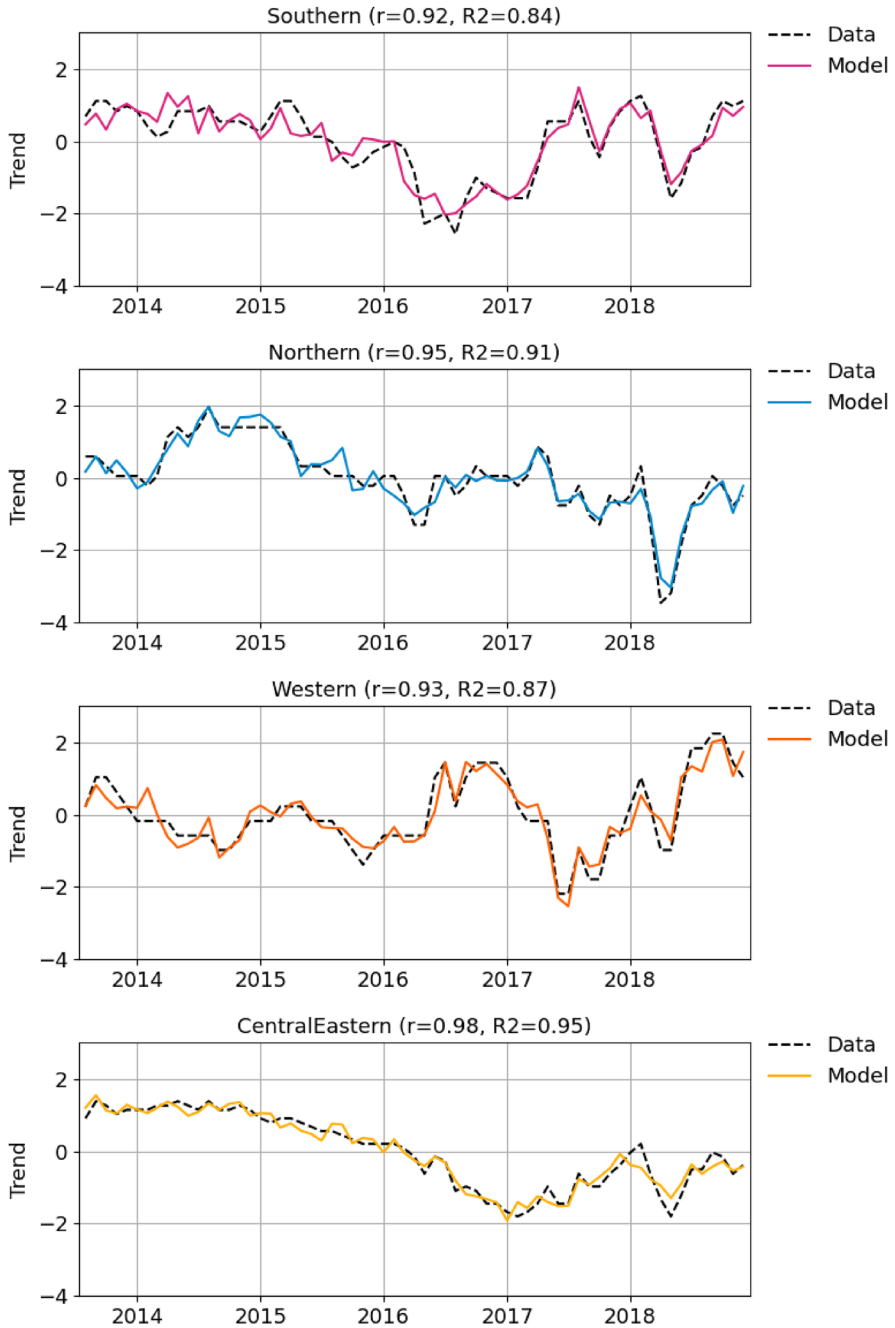


Fig. 10 Time-series comparing actual, i.e., “Data”, and predicted trends, i.e., “Model”

R^2 values in Tables 2 and 3. We note that, in general, the predicted series tend to ignore the most extreme variations of the values. This effect is most evident in the first eight months of 2018, e.g., patterns in March–April of 2018. Beyond this, the pattern of the predictions follows that of the current values in all the areas and for all the years observed.

Most variables that are maintained after the backwards procedure are statistically significant. The self regressive variables (the trend 1 and 2 years before) appear to have some effects, however coefficients are relatively low. The 1 year trend is always maintained after backwards elimination, and has always a negative value, indicating that there are trend reversals happening. The 2 year trend however is only significant for Northern and Central Eastern regions, with a positive effect in N and negative in CE.

Brexit searches—both in the UK and by regions—are relevant in most cases. Among the different time spans, the 1-month and 12-month values are those that are deemed significant almost always, with the 12-month variables also being assigned higher coefficients. This emphasises the long term effect of Brexit. In all EU regions except Western Europe, the long-term attention in the regions (*Brexit 12 months*) positively impacts the prediction of trend values, while the impact on prediction is negative when the attention comes from the UK (*UK Brexit 12 m*). That means that, as expected, a higher long-term attention to Brexit in the UK corresponds to lower *FLR* values, i.e. more persons leaving the UK. However, apparently, higher long-term attention to Brexit in the destination country corresponds to increased *FLR*. This could be due to the returner effect: a larger attention to Brexit outside the UK will cause UK citizens to return to their origin country. Short and medium-term attention (*Brexit 1/6 months*) in European regions has a negative effect on trends, i.e. as attention grows the *FLR* decrease, indicating a larger fraction of persons leaving. Furthermore, when the short-term attention comes from the UK, impact is positive. This is counter intuitive and could be an artefact of the long time required to make a decision to move, that shifts the moment in which the *FLR* starts to decrease, so that a natural decrease in Brexit attention will appear to correspond to a decrease in *FLR*. Western Europe however shows the opposite picture, with long term positive effects from searches in Western Europe and negative effects from searches in the UK. Short-term effect effects appear negative from Western Europe and positive from the UK. This could indicate a shorter time to decision, so that a large last month's attention to Brexit already causes decreases in *FLR* this month. However, on the long term we see an apparent positive relation, which could be due to continuous changes in trends.

Residence permit-related features by regions are considered inputs to the models of Northern, Southern and Western Europe. However, these are not taken into account by any final model, except Northern (6 months average). This means that Google searches related to the UK residence permit (in the regions) are not useful in predicting the main trend. Conversely, searches performed in the UK are considered in input by all base models and by most of the final ones. Coefficients are generally negative, indicating that as the search for Residence permits increase, the *FLR* decreases due to people leaving the UK. The highest effects

are from the 12 month variables, underlining again the importance of long term effects.

The impact and relevance of *visa*-related searches appears reduced in terms of the number of variables maintained in the models and coefficient ranges. Regarding the searches carried out in the regions, only the 6 and 12 month variables are significant in final models, and generally have negative coefficients. This means that persons who search for Visas take longer to move, and as the attention to visas grows the number of people leaving also grows. The searches in the UK provide a similar picture, with generally few significant variables and low coefficients, except for Southern Europe where the coefficient is larger. The effect is positive, i.e. as the visa search within the UK grows, the *FLR* grows, indicating fewer people leaving. This could be due to the fear of leaving and not being able to get a new Visa upon return.

As for the country related terms, the variables corresponding to searches for the UK in EU regions are rarely maintained and significant after backwards elimination, and have generally low coefficients. However, the searches for particular countries from the UK are significant for all regions. One-month searches have negative influence on the trend, i.e. as more persons search for EU countries, the *FLR* decreases since more persons leave. This could indicate that people search information about the country of destination short before leaving the UK. The 6-month variable, however, has a positive coefficient, probably due to periodicity in trends. Long term 12 month variables have again negative coefficients, maybe indicating initial searches when the decision to leave was first contemplated.

All in all, regression analysis shows a good relation between Google Trends data related to Brexit, including visas and residence permits, with models being able to explain very well the main component of the *FLR*, suggesting that the changes in patterns that we observe are indeed due to Brexit. While significant in models for all regions, the self regressive component is not the strongest. Generally, we see largest effects from 12-month variables, underlining the long term effects of Brexit.

Conclusions

This paper presented an analysis of two indicators aimed to understand trends in mobility between the UK and Eu27 around Brexit. We employed different data types offering a different view over the changes. Official migration data from Eurostat allowed to study general migration, Crunchbase and Scientific migration data enabled analysis of highly-skilled migration, while Air Passenger and Twitter data provided insights into general mobility. All data supported similar findings. While before Brexit the UK was an attractive destination both for general migration and highly-skilled migrants, after Brexit some EU regions showed a change in trend. The strongest inversion of trend was observed for Central and Eastern Europe, and this matches earlier findings.¹⁹ Northern Europe, however, appears to have the opposite behaviour, with a general decrease in mobility towards the UK before Brexit and a slight increase afterwards. The change in trend started already from the Brexit referendum (June 2016).

¹⁹ <https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/internationalmigration/bulletins/migrationstatisticsquarterlyreport/february2019>.

An important type of data employed here are Air Passenger Data. Their main advantage compared to other data types, including traditional data, is the time resolution and the volume. We were able to compute meaningful indicators on a monthly basis, allowing to study the seasonal patterns, impossible in other data types. While before Brexit a stable seasonality could be observed, the pattern was strongly disrupted between 2016 and 2019, for all EU regions. This could mean changes in travelling patterns of UK migrant residents, or changes in seasonal migration. In fact, the UK is one of the main recipients of seasonal migration in Europe [36]. The analysis of the link between the attention to Brexit through Google Trends and the flow log ratios has uncovered a strong relation, confirming again that the changes observed are due to the geopolitical context. From 2020, patterns changed even more due to Covid-19 interference, so we excluded this period from the regression analysis. Among disadvantages of this data type, an important one is that it includes all types of mobility, not only migration. Dissecting the different signals is not straightforward, therefore one can consider these data in conjunction with other data types that show more specific types of mobility. Our indicators are only affected in their range, with general low values, especially at the yearly level. However, we are interested in *changes in time*, and these were still clear even if absolute values of our indicators were all low.

The higher time and space resolution is also possible with social media data, Twitter in our case. However, due to selection bias, mobility flows were much more reduced in this case, therefore differences and consequently indicator values are magnified. Even if referring to a smaller and less representative population, findings from other data types were confirmed by Twitter data, even if absolute values were often different. This supports the hypothesis that Twitter and Social Media data can be employed for early monitoring of mobility and migration [34, 35, 45], however more accurate data are required for robust measurement.

Although more complete than other previous works, our study of the mobility patterns after Brexit can be improved. In particular, other data types could be investigated, including non-traditional sources such as professional social networks (e.g. LinkedIn), Facebook Advertising data, mobile phone data, that could enable an even better granularity.

Acknowledgements This work was supported by the European Union - Horizon 2020 Program through the project “HumMingBird - Enhanced migration measures from a multidimensional perspective”, Grant Agreement n.870661, and under the scheme “INFRAIA-01-2018-2019 - Integrating Activities for Advanced Communities”, Grant Agreement n.871042, “SoBigData++: European Integrated Infrastructure for Social Mining and Big Data Analytics”.

Funding Open access funding provided by Università di Pisa within the CRUI-CARE Agreement.

Data availability statement The Eurostat flows employed in this study are available on the Eurostat portal (<https://ec.europa.eu/eurostat/databrowser/explore/all/popul>). The Scientific Migration flows are available in the EMAGK dataset (<https://zenodo.org/record/5888647>) and described in detail in Pollacci [37]. This dataset also includes other variables and data types, we have extracted only UK incoming and outgoing flows for this study. Air passenger data is proprietary so we could not publish raw data, however we published the indicators computed in this study, at both monthly and yearly resolution: <https://doi.org/10.7910/DVN/AE1PKC>. Twitter (currently X) and Crunchbase data were obtained through the public Academic API, and cannot be redistributed. However, we have described the data collection methodology that can be reproduced after obtaining access to data.

Declarations

Conflict of interest On behalf of all authors, the corresponding author states that there is no Conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Ahmad-Yar, A. W., & Bircan, T. (2021). Anatomy of a misfit: International migration statistics. *Sustainability*, 13(7), 4032.
2. Ahmouda, A., Hochmair, H. H., & Cvetojevic, S. (2019). Using twitter to analyze the effect of hurricanes on human mobility patterns. *Urban Science*, 3(3), 87.
3. Aydođdu, B., Samad, H., Bai, S., Abboud, S., Gorantis, I. & Salah, A. A. (2023). 'Analyzing international airtime top-up transfers for migration and mobility', *International Journal of Data Science and Analytics* pp. 1–18.
4. Barnard, C., Butlin, S. F., & Costello, F. (2022). The changing status of european union nationals in the united kingdom following brexit: The lived experience of the european union settlement scheme. *Social & Legal Studies*, 31(3), 365–388. <https://doi.org/10.1177/09646639211032337>
5. Benson, M., Sigona, N., Zambelli, E., & Craven, C. (2022). From the state of the art to new directions in researching what brexit means for migration and migrants. *Migration Studies*, 10(2), 374–390.
6. Böhme, M. H., Gröger, A., & Stöhr, T. (2020). Searching for a better life: Predicting international migration with online search keywords. *Journal of Development Economics*, 142, 102347.
7. Bosco, C., Grubanov-Boskovic, S., Iacus, S., Minora, U., Sermi, F., & Spyrtatos, S. (2022). Data innovation in demography, migration and human mobility, Science for Policy Report KJ-NA-30907-EN-N (online), KJ-NA-30907-EN-C (print). *Luxembourg (Luxembourg)*. <https://doi.org/10.2760/027157>
8. Breinlich, H., Leromain, E., Novy, D., Sampson, T., & Usman, A. (2018). The economic effects of brexit: Evidence from the stock market. *Fiscal Studies*, 39(4), 581–623.
9. Campos, R. G. & Timini, J. (2019). 'An estimation of the effects of brexit on trade and migration', *Banco de Espana Occasional Paper* (1912).
10. Chanda, R., & Betai, N. V. (2021). Implications of brexit for skilled migration from india to the uk. *Foreign Trade Review*, 56(3), 289–300. <https://doi.org/10.1177/00157325211012207>
11. Chi, G., Lin, F., Chi, G., & Blumenstock, J. (2020). A general approach to detecting migration events in digital trace data. *PLoS one*, 15(10), e0239408.
12. Eurostat (n.d.a). 'Emigration by age group, sex and citizenship', <https://bit.ly/ESTATOutflowCit>. Accessed: February 2022.
13. Eurostat (n.d.b). 'Emigration by age group, sex and country of next usual residence', <https://bit.ly/ESTATOutflowRes>. Accessed: February 2022.
14. Eurostat (n.d.c). 'Immigration by age group, sex and citizenship', <https://bit.ly/ESTATInflowCit>. Accessed: February 2022.
15. Eurostat (n.d.d). 'Immigration by age group, sex and country of previous residence', <https://bit.ly/ESTATInflowRes>. Accessed: February 2022.
16. Eurostat (n.d.e). 'International migration statistics', https://ec.europa.eu/eurostat/cache/metadata/en/migr_imm_i_esms.htm. Accessed: February 2022.
17. Fantazzini, D., Pushchelenko, J., Mironenkov, A., & Kurbatskii, A. (2021). Forecasting internal migration in russia using google trends: Evidence from moscow and saint petersburg. *Forecasting*, 3(4), 774–804.

18. Färber, M. (2019). The Microsoft Academic Knowledge Graph: A Linked Data Source with 8 Billion Triples of Scholarly Data, in 'Proceedings of the 18th International Semantic Web Conference', ISWC'19, pp. 113–129. https://doi.org/10.1007/978-3-030-30796-7_8
19. Färber, M. & Ao, L. (2022). 'The microsoft academic knowledge graph enhanced: Author name disambiguation, publication classification, and embeddings', *Quantitative Science Studies* pp. 1–44.
20. Fiorio, L., Abel, G., Cai, J., Zagheni, E., Weber, I. & Vinué, G. (2017). Using twitter data to estimate the relationship between short-term mobility and long-term migration, In: 'Proceedings of the 2017 ACM on web science conference', pp. 103–110.
21. Gabrielli, L., Deutschmann, E., Natale, F., Recchi, E., & Vespe, M. (2019). Dissecting global air traffic data to discern different types and trends of transnational human mobility. *EPJ Data Science*, 8(1), 26. <https://doi.org/10.1140/epjds/s13688-019-0204-x>
22. Gabrielli, L., Deutschmann, E., Natale, F., Recchi, E., & Vespe, M. (2019). Dissecting global air traffic data to discern different types and trends of transnational human mobility. *EPJ Data Science*, 8(1), 26.
23. Iacus, S. M., Natale, F., Santamaria, C., Spyrtos, S., & Vespe, M. (2020). Estimating and projecting air passenger traffic during the covid-19 coronavirus outbreak and its socio-economic impact. *Safety Science*, 129, 104791.
24. Kilkey, M., Piekut, A. & Ryan, L. (2020). 'Brexit and beyond: Transforming mobility and immobility', *Central and Eastern European Migration Review* pp. 5–12.
25. Kim, J., Pollacci, L., Rossetti, G., Sirbu, A., Giannotti, F., & Pedreschi, D. (2022). Twitter Data for Migration Studies, in 'Data Science for Migration and Mobility'. *British Academy*. <https://doi.org/10.5871/bacad/9780197267103.003.0008>
26. Kim, J., Sirbu, A., Giannotti, F. & Gabrielli, L. (2020). Digital footprints of international migration on twitter, In: 'International Symposium on Intelligent Data Analysis', Springer, pp. 274–286.
27. Lai, S., Sorichetta, A., Steele, J., Ruktanonchai, C. W., Cunningham, A. D., Rogers, G., Koper, P., Woods, D., Bondarenko, M., Ruktanonchai, N. W., et al. (2022). Global holiday datasets for understanding seasonal human mobility and population dynamics. *Scientific Data*, 9(1), 1–13.
28. Li, P., Lu, H., Kanhabua, N., Zhao, S., & Pan, G. (2018). Location inference for non-geotagged tweets in user timelines. *IEEE Transactions on Knowledge and Data Engineering*, 31(6), 1150–1165.
29. Luthra, R. R. (2021). Mitigating the hostile environment: the role of the workplace in eu migrant experience of brexit. *Journal of Ethnic and Migration Studies*, 47(1), 190–207. <https://doi.org/10.1080/1369183X.2020.1726173>
30. Markova, E., & King, R. (2021). Leave or remain? the post-brexit (im) mobility intentions of bulgarians in the united kingdom. *European Urban and Regional Studies*, 28(1), 58–65.
31. Martín, Y., Cutter, S. L., Li, Z., Emrich, C. T., & Mitchell, J. T. (2020). Using geotagged tweets to track population movements to and from puerto rico after hurricane maria. *Population and Environment*, 42(1), 4–27.
32. Mayer, T., Vicard, V., & Zignago, S. (2019). The cost of non-europe, revisited. *Economic Policy*, 34(98), 145–199.
33. McKinney, W., Perktold, J. & Seabold, S. (2011). 'Time series analysis in python with statsmodels', *Jarrodmillman. Com* pp. 96–102.
34. Minora, U., Belmonte, M., Bosco, C., Johnston, D., Giraudy, E., Iacus, S. & Sermi, F. (2022). 'Migration patterns, friendship networks, and the diaspora: the potential of facebook's social connectedness index to anticipate migration flows induced by russia's invasion of ukraine in the european union', *arXiv preprint*. [arXiv:2209.01833](https://arxiv.org/abs/2209.01833) .
35. Minora, U., Bosco, C., Iacus, S. M., Grubanov-Boskovic, S., Sermi, F., & Spyrtos, S. (2022). The potential of facebook advertising data for understanding flows of people from ukraine to the european union. *EPJ Data Science*, 11(1), 59.
36. OECD. (2010). *International migration outlook*. SOPEMI Report: Technical report.
37. Pollacci, L. (2022). 'Emakg: An enhanced version of the microsoft academic knowledge graph', *arXiv preprint*. [arXiv:2203.09159](https://arxiv.org/abs/2203.09159) .
38. Portes, J. (2022). Immigration and the UK economy after Brexit. *Oxford Review of Economic Policy*, 38(1), 82–96. <https://doi.org/10.1093/oxrep/grab045>
39. Portes, J., & Forte, G. (2017). The economic impact of brexit-induced reductions in migration. *Oxford Review of Economic Policy*, 33(suppl-1), S31–S44.
40. Priedhorsky, R., Culotta, A. & Del Valle, S. Y. (2014). Inferring the origin locations of tweets with quantitative confidence, In 'Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing', pp. 1523–1536.

41. Sabre. (2021). 'Market intelligence, global demand data', <https://www.sabre.com/products/market-intelligence/>. Accessed: 2021-11-15.
42. Salah, A. A., et al. (2022). Can big data deliver its promises in migration research? *International Migration*, 60(2), 252–255.
43. Simionescu, M., Bilan, Y., Smrčka, L. & Vincúrová, Z. (2017). 'The impact of brexit on labour migration and labour markets in the united kingdom and the eu', *Terra Economicus*.
44. Simionescu, M., Strielkowski, W., & Kalyugina, S. (2017). The impact of brexit on labour migration and labour markets in the united kingdom and the eu. *Terra Economicus*, 15(1), 148–156.
45. Sîrbu, A., Andrienko, G., Andrienko, N., Boldrini, C., Conti, M., Giannotti, F., Guidotti, R., Bertoli, S., Kim, J., Muntean, C. I., et al. (2021). Human migration: the big data perspective. *International Journal of Data Science and Analytics*, 11(4), 341–360.
46. Spyrtatos, S., Vespe, M., Natale, F., Weber, I., Zagheni, E. & Rango, M. (2018). 'Migration data using social media: a european perspective'.
47. Stawarz, N. & Witte, N. (2021). 'Unintended repercussions of brexit? the departure of skilled germans from britain after the referendum'.
48. Sumption, M. & Walsh, P. W. (2022). briefing - EU migration to and from the UK, Technical report, Oxford University. <https://migrationobservatory.ox.ac.uk/wp-content/uploads/2020/10/MigObs-Briefing-EU-migration-to-and-from-the-UK.pdf>
49. Trabka, A., & Pustulka, P. (2020). Bees & butterflies: Polish migrants' social anchoring, mobility and risks post-brexit. *Journal of Ethnic and Migration Studies*, 46(13), 2664–2681.
50. Vargas-Silva, C. (2016). EU Migration to and from the UK After Brexit. *Intereconomics: Review of European Economic Policy*, 51(5), 251–255.
51. Veillon, L. (2021). 'How to bulk download data from google trends using python'. <https://medium.com/start-with-data-science/how-to-bulk-download-data-from-google-trends-using-python-69646fd46351>.
52. Yildiz, D., Munson, J., Vitali, A., Tinati, R., & Holland, J. A. (2017). Using twitter data for demographic research. *Demographic Research*, 37, 1477–1514.
53. Zagheni, E., Weber, I. & Gummadi, K. (2017). 'Leveraging facebook's advertising platform to monitor stocks of migrants', *Population and Development Review* pp. 721–734.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Alina Sîrbu¹  · Diletta Goglia^{1,2} · Jisu Kim³ · Paul Maximilian Magos¹ · Laura Pollacci¹ · Spyridon Spyrtatos⁴ · Giulio Rossetti⁵ · Stefano Maria Iacus⁶

✉ Alina Sîrbu
alina.sirbu@unipi.it

¹ Department of Computer Science, University of Pisa, Pisa, Italy

² Department of Information Technology, Uppsala University, Uppsala, Sweden

³ Max Planck Institute for Demographic Research, Rostock, Germany

⁴ European Union Intellectual Property Office, Alicante, Spain

⁵ Institute of Information Science and Technologies "A. Faedo" (ISTI), National Research Council (CNR), Pisa, Italy

⁶ Institute for Quantitative Social Science, Harvard University, Cambridge, USA