UNIVERSITÀ COMMERCIALE ”LUIGI BOCCONI”
PhD SCHOOL

PhD program in: Public Policy and Administration
Cycle: 32^{nd}
Disciplinary Field (code): SPS/07

Digital and Computational Approaches to Migration Studies: 3 Essays

Advisor: Professor Francesco C. Billari
Co-Advisor: Professor Emilio Zagheni

PhD Thesis by:
Aslı Ebru Şanlıtürk
ID number: *3028737*

Year 2021

Abstract

This dissertation aims to contribute to the literature on computational social sciences and presents three essays in migration studies and demography, using digital data and computational methods. The first essay focuses on visual comparison of migration patterns using Turkey as a case study. The internal migration patterns in Turkey are compared with the settlement patterns of Syrians under temporary protection in Turkey, while questioning whether there is a possibility for replacement migration policies. The second essay also uses the case of Syrians under temporary protection in Turkey and contributes to the literature on nowcasting & forecasting based on digital data by following the mobility patterns of Syrians inside Turkey using online search data from Google Trends. The third essay contributes to the literature on high-skilled migration and the use of bibliometric data. The essay uses the Brexit decision in 2016 and the academic environment in the United Kingdom as a case study and monitors the change that occurred in the in- and out-migration patterns of researchers with respect to the UK, before and after the Brexit referendum.

Keywords: Computational Social Science, Migration, Refugees, Turkey, Big Data, Brexit

Acknowledgements

This dissertation has been enabled by the invaluable support and guidance of Prof. Francesco C. Billari and Prof. Emilio Zagheni, to whom I am deeply grateful. My sincere gratitude also to the members of supervisory committee Prof. Carlo Devillanova and Prof. Nicoletta Balbo for their guidance and input.

I would like to express my gratitude to Max Planck Institute of Demographic Research (MPIDR), for granting me access to their academic infrastructure and database and for their collaboration in the Chapter 4 of this dissertation. I would like to especially thank Dr. Samin Aref for his collaboration and help and the members of the Laboratory of Digital and Computational Demography for their support.

Special thanks also to the members of the DisCont (Discontinuities in Household and Family Formation) research group at Bocconi University for their support and feedback.

Last but not least, I would like to thank my family as well as my family away from home. I cannot imagine reaching where I am today without the never-ending love and support of my parents. I also cannot imagine being able to finish the entire PhD journey and this dissertation without the company and support of my PhD cohort and flatmates, that became my family away from home. I thank each and every member of my PhD family (including our quadruped coworker) for their constant help, understanding and solidarity, for keeping me sane during the hardest and most challenging times of these five years, for making my 30th birthday the most memorable one I ever had, and for all the fun, hardship, joy and tears we shared together that made me a better person.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Migration is a phenomena that can be broadly defined as “movement of persons away from their place of usual residence, either across an international border or within a State (IOM, 2019)”. Beyond this general definition, migration with all its types, processes, underlying reasons and consequences, is an issue of interest in various disciplines of social sciences. Perceived from different perspectives and dimensions, migration touches, changes and shapes individuals (micro level), societies and states (macro level), global systems and relations (meso level) and policies of a wide array from enabling to restrictive (policy level).

The attempts to develop a theoretical framework for migration continue since the famous study of Ernst Georg Ravenstein (1885) defining the laws of migration. It is not possible to develop an all-embracing single theory for a phenomena as migration, yet the fundamental assumptions on the laws of migration that Ravenstein had laid and Everett Lee (1966) had revised still constitute the starting point in migration studies. Migration is considered to be influenced by four main factors; factors related to the origin, factors related to the destination, intervening obstacles and personal factors (Lee, 1966, p.50). The factors related to the origin and destination, fostering migration from one place to another, are commonly referred as push and pull factors respectively (Passaris, 1989). Factors such as the geographic distance, geographic irregularities, policy-based migration barriers, differences of culture and language affecting the accessibility constitute the intervening obstacles. Personal factors are individual traits that alter the weight of local and society-level conditions for different people, which may mitigate or increase the intervening obstacles.

However, relying only on the effects of push and pull factors to understand the underlying reasons of migration would be misleading. The dichotomies that are used to explain migration flows such as push vs. pull, labor-abundant vs. labor-scarce, underdeveloped vs. developed shadow the depth of the migration processes and often neglect the assumption of currents and counter-currents, hence the circularity of migration. In fact, literature shows that migration patterns follow a similar pattern of demographic transition, in terms of its relationship with development. In contrast to the initial theories on international migration, development accelerates migration flows until an equilibrium is reached, which is referred as the migration transition (De Haas, 2007, 2010; Clemens, 2014). Furthermore, with higher levels of development and the accompanying increase in the education levels (*capabilities*) paves the way for larger volumes of high-skilled migration, as the capabilities and aspirations of people increase. Migration of high-skilled people is also subject to a dichotomous conceptualization as brain drain vs. brain gain, however, recent studies propose that the high-skilled migrants are under the influence of different factors concerning their migration patterns (Mahroum, 2000) and high-skilled migration should be considered as a pattern of brain circulation rather than a one-directional flow (Saxenian, 2005).

More than a century after Ravenstein's work on laws of migration based on the population data in United Kingdom, a new source of data appeared for the social scientists. The emergence of online generated digital and big data acquired attention as a source for social science research and ever since the literature comprising of the analyses and interpretation of the digital footprints of individuals began to grow. Digital data and computational methods came to be considered as especially helpful, in cases where formal data, such as official registries, census data or national surveys, are missing or lacking (Alburez-Gutierrez et al., 2019), where data collection is burdensome or limited due to accessibility, sensitivity or confidentiality issues (Billari et al., 2020), where the use of digital devices and social media is widespread and where the aim is to analyse the immediate (even simultaneous) results of policies before the long traditional data collection process is complete (Ginsberg et al., 2009). The latter motivation to use digital data especially gathered attention following the breakout of the COVID-19 pandemic (Brodeur et al., 2020; Wilde et al., 2020).

In the field of migration studies, migration networks theory posits that the existence of established migrant communities, historical migrants and diasporas contribute to the incentives to migrate (Taylor, 1986; Massey and España, 1987; Gurak and Caces, 1992). Due to its similarity with conceptualizing online networks, migration networks approach addressed the potential of new ICT technologies and social media platforms first, to eliminate accessibility and information related barriers (Diminescu, 2008) and facilitate migration (Dekker and Engbersen, 2014). Acknowledging that migrants tend to use ICT technologies and social media platforms effectively (Diminescu, 2008), handed migration studies a new source of data collection on migrants, where the traditional methods may fail to provide sufficient information (Zagheni et al., 2017b).

The use of digital data in migration studies also proved useful for the studies on refugees, as the online platforms closed the information gap for refugees, they also closed gap of missing data for researchers. Studies show that Syrian refugees consider smart phone ownership and use of social media as essential as food and shelter, as they rely on these means to navigate in an unfamiliar environment as well as keep contact with their friends and families (Gillespie et al., 2018; Dekker et al., 2018). Widespread use of mobile phones and social media platforms by the Syrian refugees also enable research based on their digital footprints, as seen in the case of *Data* *for Refugees - D4R Challenge* in Turkey (Salah et al., 2018).

The aim of this dissertation is to contribute to the use of digital data and computational approaches in demography and migration studies. Chapters 2, 3 and 4 in this dissertation focus on the use of digital data and computational methods in migration studies. More specifically the essays concentrate on two different types of migration, chapters 2 and 3 on forced migration & refugees and the chapter 4 on high-skilled migration. Chapters 2 and 3 are based on the case of Syrian refugees in Turkey while in Chapter 4 the focus is on the internationally mobile scientists as subjects of high-skilled migration rather than a single context.

Chapter 2 in this dissertation uses the digital and computational methods in data visualisation to interpret and compare internal migration flows and settlement of Syrian refugees. In that respect, the chapter uses the digital and computational methods for the visualisation and interpretation of the data as well as the data collection by Google Maps Distance API, while still relying on official sources for the main data and diverging from chapters 3 and 4. The chapter then seeks an answer to the question, to what extent the migration behaviour of Syrians under temporary protection in Turkey is correlated with the internal migration behaviour of Turkish citizens inside Turkey, using quantitative methods.

Chapter 3 of the dissertation also focuses on the case of Syrians under temporary protection in Turkey however, this chapter introduces the use of digital and computational methods also into the data collection process. In chapter 3, data obtained from Google Trends are used as a tool to analyse the change in Syrian refugee stock across provinces in Turkey. Thus, digital data and computational methods are used both in data collection and analysis parts of the chapter. In data collection, these methods contributed not only to the retrieval of Google Trends data but also recreate the official registry dataset, through Wayback Machine and WebCite. The aim is to understand whether the search frequency for a province name may be associated with the refugee stock of the said province and whether the Google Trends data may be used for forecasting the refugee flow to different provinces, in a way to help the migration management policies.

In both chapters, the Syrian refugees are referred as 'Syrians under temporary protection' (Syrians uTP) referring to the legal status of Syrian citizens in Turkey, i.e. temporary protection. It should be underlined that this status differs from the general meaning of refugee and asylum seeker in international law. In Turkey, this status is created for and only applicable to the case of Syrians after the Syrian civil war. Obtaining the temporary protection status requires official registration by authorities, allowing them legal protection and placing them out of the scope of irregular migration. Underlining this issue in the beginning is important as the migration/mobility patterns of Syrians under TP, may show differences to of other migrant groups and unregistered migrants. I acknowledge that the issue of unregistered migrants that is out of the scope of this paper and welcome any research that can be made in this field, although current data on irregular migration is not sufficient.

Chapter 4 of this dissertation focuses on the international mobility of scientists as a case of high-skilled migration. It uses the data obtained from the Scopus database, that includes the publication information (title, journal name and DOI), individual author code and author affiliation in terms of both institution and country. The aim of chapter 4 is to use the big data retrieved from Scopus database to analyse the effects of United Kingdom's withdrawal from the EU (known as Brexit) and the uncertainty it created during the period of transition (2016-2020).

Chapter 2

Refugee Inflows and Sustainable Migration Policies: A Case Study in Turkey

Ebru Sanliturk [1]

1. Bocconi University

The recent refugee crisis in Europe, also referred as the Mediterranean refugee crisis or Syrian refugee crisis, that emerged following the break out of the civil war in Syria in 2011 and reached its peak in 2015, is an important issue of concern for policy-makers and researchers. As the economic and social burden of the refugee influx increases, especially in countries that now host a large population of refugees, understanding the patterns of refugee flows have become more important than ever in order to develop sustainable policies for refugee mobility.

This study focuses on the case of Turkey as a host country and attempts to shed some light to the settlement patterns of Syrians in Turkey under the temporary protection status (hereafter Syrians uTP), to contribute to the development of sustainable migration policies. Selection of Turkey as the case study relies on two reasons. First, the size of the refugee shock Turkey had experienced is much greater than the shock experienced by most of Mediterranean countries in terms of actual numbers, salience and economic and social burden. The shock of this refugee inflow to Turkey is only comparable with Lebanon, hosting the highest number of refugees in the world in relative numbers, while Turkey hosts the highest number of refugees in absolute terms (UNHCR, 2018). Second, while most EU states host Syrian refugees in special migration reception centres following their arrival, this is not possible in the case of Turkey. Of the 3.6 million Syrians uTP in Turkey, less than 60,000 live in refugee camps [1]. Thus, Syrians uTP are a part of urban life and their mobility can be tracked through the temporary protection registration system. As the Syrian refugee crisis is ongoing for 8 years now and estimated to continue, also bearing in mind that not every person currently under temporary protection in

[1] According to the official statistics provided by the Directorate General of Migration Management in Turkey.

Turkey will return to Syria, even if peace conditions would be achieved, the necessity of long-term policy-making is evident.

This paper focuses only on the Syrians uTP in Turkey, using the temporary protection registries, and their settlement patterns, applying a comparative perspective with the internal migration patterns of the local population. I acknowledge that other migrant groups, living in Turkey due to regular (registered) or irregular migration may have similar or different settlement patterns, yet this topic is outside the scope of this study. Further research in this field may reveal the similarities and differences between different migrant groups, although current data on irregular migration is not as detailed. The temporary protection status differs from the general meaning of refugee and asylum seeker and is only applicable to the case of Syrian citizens in Turkey after the Syrian civil war.

This paper aims to contribute to the migration studies in the context of Turkey in two ways. The first section of the study, presents a comparison between the internal migration patterns in Turkey and mobility of Syrians uTP in Turkey using data visualization tools. Visual analysis of migration movements is witnessing a growing interest in the literature with inspiring studies such as the adoption of genomic visualisation methods to map bilateral international migration flows (Abel and Sander, 2014) and dendrochronology (tree-ring dating) visualisation methods to map the US immigration patterns (Cruz et al., 2019). In this study I will follow the approach of Abel & Sander (2014) and to the best of my knowledge this method has not been used in the context of Turkey before, nor in a manner to graph the comparison between local and migrant/refugee populations. The main objective of using this method is to understand whether there is a visible convergence and/or divergence between the internal migration patterns and settlement patterns of Syrians uTP, and how these patterns can be interpreted. In the second section of the study, an empirical analysis is carried out to shed light to what drives the settlement patterns of Syrians uTP in Turkey and what the policy implications of these patterns could be. Each of these parts include a data and methodology part and a part for results. Using both data visualization techniques and empirical analysis results, I inquire whether replacement migration may be a policy option for Turkey and Syrians under temporary protection. Thus, in conclusion, the results of the two sections of the study are discussed within the framework of fundamental migration theories and replacement migration.

2.1 Background and Literature Review

2.1.1 Background

Migration, both international and internal, has always been an important issue for Turkey. Starting with the population exchange agreement with Greece in 1923, Turkey experienced migrant flows, where the migrants mostly had ethnic or cultural ties with Turkey. In the last four decades international migration flows based on ethnic and cultural ties continued with migrants from Iraq, Bulgaria, Bosnia, Kosovo and Macedonia[2].

Internal migration in Turkey started to become a prominent issue starting from the 1950s. Internal migration in Turkey, just as the migration trends in the world, is often a migration from rural to urban areas or between urban areas. Large metropolitan areas emerged as result of this exodus to urban areas, most striking example of which is Istanbul, where the population increased form 1.5 million in 1955 to 15.5 million in 2020. The findings of Eryurt and Koç (2015) suggest that Turkey is a migration country, as 63.3% of the population over the age of 17 have migrated at least once in their lifetime. Overall, urban to urban migration accounts for 29.3% of internal migration trends, while rural to urban follows with 24.9% (Eryurt and Koç, 2015, p.499). It should also be noted that internal migration trends and its reasons in Turkey may vary among ethnic groups. As result of the armed conflict with the Kurdish separatists, Kurds in Turkey, especially during the 1990s, were internally displaced or forced to migrate due to security reasons, from rural areas in the southeast to urban areas in the west. This increased population movement of Kurds in Turkey is also apparent in the Survey on Migration and Internally Displaced Persons in Turkey, 2005, as the share of ever migrating is 71.7% among Kurds and 61.1% among Turks. About 12% of the Kurds stated to have migrated due to security reasons and while Kurds are observed to have a rural to urban migration trend, Turks are observed to migrate among urban areas (Eryurt and Koç, 2015, p.499).

The breakout of violence in 2011 in Syria turned into a civil war within months and escalated to a point where the gravity of the refugee crisis extended beyond the immediate region and neighbouring countries. Since 2011, Turkey applied an open-door policy to Syrian refugees, at first considering that civil war would end soon and repatriation would be possible after the end of the civil war. As the gravity of the crisis and the influx of refugees increased in 2014 and 2015, the Turkish government accepted to become a part of Syrian Regional Response Plan (SRRP) in cooperation with the UNCHR (Kirişci and Ferris, 2015) as well as signed the controversial readmission agreement with the EU (known as the EU-Turkey Deal). Per the EU-Turkey Deal, Turkey agreed to strengthen its capabilities to register, accommodate and facilitate Syrians while the EU commits to provide financial assistance to Turkey.

As of 28 October 2020, Turkey hosts 3,627,991 registered Syrians uTP according to data published by the DGMM in Turkey. Syrians uTP outside the camps can move between provinces due to family, health, work and education purposes, yet they need to register their new address in order to continue to receive assistance. Currently there are seven shelter centres for Syrians uTP, which host 59,427 people in 5 provinces. These figures leave 3,568,534 people living in urban areas (of Migration Management, 2020).

[2]Figures are received from the Directorate General of Migration Management.

2.1.2 Literature Review

Internal migration in Turkey has been elaborated by researchers mainly focusing on two issues; its economic implications (development and inequality) and social implications. Regarding the economic implications, Çıracı and Atalık (1993) present an empirical analysis, showing that the push effect originates from regional differentiation and the most significant factor affecting the net change in migration is GNP per capita. In another study on the effects of per capita income on net migration rate, Yamak and Yamak (1999) show that the pull-effect appears to be more important for internal migration in Turkey, as high income of migrant receiving provinces is observed to be more significant than the low income of provinces of out-migration. Gezici and Keskin (2005) also provide evidence that per capita income and job opportunities are the main drivers of internal migration, the direction of migration is often to the west and being located at the coast increases the pull effect of provinces, except for the Black Sea region. Furthermore, Gezici and Hewings (2004) conduct a spatial analysis to examine whether there is convergence between the regional inequalities over time and conclude that no significant convergence can be observed.

Looking at both economic and social aspects of internal migration, Gökhan and Filiztekin (2008) provide evidence to the intuitional assumption that income-seeking migrants are younger and have higher education levels in comparison to people migrating for other reasons. They also find a gender variation and argue that, although one gender cannot be defined as the dominant income-seeker over the other, for females the variation suggests a certain degree of migration-dependency on males (2008, p. 25).

The social implications of migration are also examined from the perspective of ethnicity, considering the ethnic variation in migration in Turkey mentioned above, despite the substantial lack of data. In social sciences literature, the ethnic distribution in Turkey is often measured by the mother tongue based on the results of population censuses. The 1965 population census is the last one that reports the results of the distribution of different mother tongues, allowing for consistent estimation of ethnic minorities in Turkey [3]. Recently, Turkish Demographic Health Survey data replaced the population censuses as a source for ethnicity estimation, as it includes the mother tongue question starting from 1998. Despite the challenges of data availability, studies on ethnicity and migration increased in the last decades. Analysis of the significant difference in fertility rates and fertility decline between Turks and Kurds across and within regions (Koc et al., 2008) and language shifts due to internal migration, where bilingualism and even the number of monolingual-Turkish speakers increased among the Kurds migrating to West, especially if they receive higher education (Zeyneloglu et al., 2016) are some prominent examples in this field.

[3] Until the 1990 population census, the relevant question had remained as part of the questionnaire but results were not reported. Eventually the question had been removed from census questionnaire form in 1990 and the questionnaire-based census data has been abandoned in 2007 to be replaced by Address Based Population Registration System (ABPRS) for population statistics.

Understanding the settlement patterns of Syrians uTP is important in the case of Turkey, as it is a unique case that hardly fits into any strict theoretical approach. The migration of Syrians uTP does not fit into traditional refugee understanding, considering that only less than 5% of the refugees live in camp environments. By a great majority, Syrians uTP in Turkey are living in urban areas, seeking jobs. However, it cannot exactly fit into the framework of labour migration by De Haas (2005, p. 1271), for example, who argues that the depiction of unprecedented mass migration flows is flawed in comparison to the migrations of the former century, and labour migration is not about absolute poverty but relative deprivation due to global inequalities. Although the settlement of Syrians uTP in urban areas may resemble the labour-motivated migration patterns from developing countries, there is an absolute deprivation dimension as many Syrians uTP had lost their homes and families. Furthermore, Turkey, as a developing country itself with a relatively young population, is not comparable to the developed countries receiving labour migration. Even though the refugee influx came as an exogenous shock, it is apparent that it cannot be addressed via temporary protection policies. In that sense, Syrians in Turkey are neither immigrants who came as job-seekers nor long-time guests who will surely return eventually.

In line with that thought, Disaster and Emergency Management Authority (AFAD) in Turkey conducted an extensive survey research to assess the demographic composition, life conditions and expectancy of Syrians uTP. This study shows that 21.5% of survey respondents in camps and 15.6% of survey respondents living in urban areas stated that they never intend to return to Syria, even if peace conditions are achieved[4] (AFAD, 2017, p. 110). These results alone underline the necessity of long-term planning and integration policies.

In this study I hypothesize that the theory of replacement migration, proposed by the UN Population Division (2001) may offer a solution to the settlement of Syrians uTP in Turkey. Replacement migration proposes in essence, that international labour migration can be a solution to developed countries that decline in (working age) population. In order to be able to suggest replacement migration as a sound policy option, I first aim to understand the convergences and divergences of migration patterns of both locals and Syrians uTP in Turkey. If the settlement patterns and motivations of Syrians uTP can be assessed, policies to encourage their migration to the rather depopulated areas of Turkey, due to out-migration, might follow. In a similar understanding, Bansak et al. (2018) suggest to improve refugee integration by using an algorithm to assign incoming refugees to locations, that would better suit their background, educational, linguistic and cultural characteristics. Applying the suggested method in the contexts of United States and Switzerland, they claim that employment chances of refugees can be increased by 41% and 73% respectively (Bansak et al., 2018, p. 3).

[4]Survey respondents are asked to select one of the following options: “I never intend to go back”, “No opinion”, “I want to go back as soon as possible”, “I want to go back when the conflict in Syria is over”, “I want to go back when the conflict in my hometown is over” and “I want to go back when the regime/government changes”. Translations are my own.

2.2 Data Visualisation

2.2.1 Data and Methods

Data used in this section is obtained from the Turkish Statistical Institute (hereafter TurkStat) database and (Ministry of Interior) Directorate General of Migration Management (hereafter DGMM) registries. The internal migration data published by the TurkStat is very detailed and reports the number of people moving from one place to another at NUTS-I (regions), NUTS-II (sub-regions) and NUTS-III (provinces) levels. For the convenience of interpreting the charts, NUTS-I level data is used for internal migration. The internal migration data by TurkStat is based on the Address Based Population Registration System (hereafter ABPRS). ABPRS covers the Turkish citizens as well as foreign nationals and former Turkish citizens, with residence permits. The foreign nationals with work/residence permits do not include Syrians uTP, as the temporary protection status is not part of the definition of de-jure population based on residence (Institute, 2020). In addition to the annual in- and out-migration data, ABPRS also includes data on migrant stocks at NUTS-III (province) level. Migrant stocks are reported in two ways, province of residence by birthplace (province or district level) and province of residence by province of family registry. Moving the family registry to the place of residence is a bureaucratic procedure that many people choose not to, as the registry of current residence alone is sufficient to access necessary public services, to vote etc. Considering this aspect of the family registry system and the automatic registry of the birthplace, I used the province of residence by province of birthplace data, to visualize the migrant stocks of provinces in 2019.

The NUTS classification in Turkey is shown in Table I in the Appendix (2.5). Data obtained from the DGMM shows the distribution of refugees at province level, however, to better match the refugee settlement pattern and internal migration flows, this data is also clustered at NUTS-I level and used to visualize the settlement patterns of Syrians uTP in Turkey.

Two different methods are employed to visualize the internal migration and refugee settlements in Turkey. Considering that internal migration needs to demonstrate both in- and out-migration between NUTS-I regions, the Chord diagram is selected as the suitable method. The method used for the internal migration patterns in this study is first used in migration studies by Guy Abel and Nikola Sander (2014), to map the worldwide international migration flow. Using this innovative approach, Abel and Sander demonstrate the international migration between 1990 and 2010 for 192 countries, clustered at 15 regions. Quantifying international migration flows is considered as a useful method to reach beyond methodological boundaries and understand the causes and consequences of worldwide migration flows (Abel and Sander, 2014). Following the introduction of the circular plots and the Chord diagram into migration studies, the approach is adopted also for the analyses of internal migration flows. Chord diagram is used to plot and demonstrate the dynamics of internal migration in Australia (Charles-Edwards et al., 2015), China (Qi et al., 2017), South Korea (Abel and Heo, 2018) and Brazil (Baptista et al., 2018).

As the data on Syrians uTP allows only for an analysis of one directional flow from Syria to the current place settlement rather than a bilateral one among the NUTS-I regions, Sankey diagram is chosen as the method to employ. Using Sankey diagrams, I visualise both the dispersion of Syrians uTP and the change in the number of Syrians uTP across NUTS-I regions. In both cases, the graphs and charts are produced using R software and Chord diagram (Gu et al., 2014), Sankey diagram packages. The distribution of Syrians uTP across NUTS-I regions is mapped and included in the Appendix 2.5. Finally, following the example of Faggian and Franklin (2014), the net migration of Turkish citizens and Syrians uTP are compared at NUTS-I regions between 2016-2019 to contribute to the discussion of convergences and divergences of their respective migration patterns.

To facilitate the interpretation of both internal migration patterns and the settlement (and possibly re-settlement) patterns of Syrians under TP between NUTS-I regions in Turkey, a NUTS-I map of Turkey is presented in Figure 2.12 in the Appendix 2.5. The circular demonstration of internal migration patterns among the NUTS-I clusters of Turkey is implemented from different perspectives. For the last available year of data (2019), visualizations of both dynamic annual bilateral migration flows and static migration stocks are created. For the rest of the available years of data (form 2008 to 2018) only the bilateral migration flows are illustrated. Furthermore, to observe if and how the year-based trends would accumulate, I aggregated the annual data for the last decade (Figure 2.5-2.6). It is important that Chord diagram enables the visualisation, better perception and interpretation of the comparison between in- and out-migration across regions and to observe the streams as well as the counter-streams of inter-regional migration. However, to simplify, I also calculated the net migration on the 10-year dataset and illustrated the cumulative net migration plot (Figure 2.7). Considering the complexity of the circular flow charts and to simplify the interpretation, a second copy of each plot is made, to only show the top 10 largest links (Figure 2.8).

2.2.2 Results

The chord diagrams seem complex at first sight, due to the confusion created by multiple lines indicating migration flows in various directions. Yet the circular visualization also makes it easier to observe persistent migration patterns over the years. In the plots presented in this section as well as in the Appendix 2.5, the most dominant player in the game, the region possessing the largest volume of links with other regions, is placed on the top of the plot. In the case of Turkey, this region is always Istanbul, denoted by TR1. Each sector in the diagram shows one NUTS-I region with a specific color and labelled just outside the grid. Out-migration from each NUTS-I region is demonstrated with the same colour as the region's grid border. The streams that leave the region stand one unit farther than the grid and their destination is indicated in the multicolored line between the link and grid border. In contrast, in-migration links to NUTS-I regions have arrowheads to indicate the direction and they stand one unit closer to the destination grid border, next to the respective out-migration links.

The Chord diagrams plotting the bilateral internal migration flows and current migration stocks in 2019 are presented below in Figures 2.1-2.2 and 2.3-2.4 respectively. All values on grid axes are to be considered in thousands. The circular plots of bilateral migration flows as well as the copies showing the top ten largest streams for the period 2008-2018 can be viewed in the Appendix 2.5, figures 2.13-2.34.

Consistent with the theoretical and empirical work, Istanbul (TR1) emerges as the giant dominating all charts both in terms of in- and out-migration. One can observe the gradual increase of out-migration in Istanbul over the years, and also how the city became a region of out-migration, in the last two years, after decades of constantly positive net migration rate. While interpreting the internal migration flows in Turkey, there are two issues to keep in mind in addition to the economic reasons, proximity effect and migration of Kurds. It can be seen that there is a persistent migration pattern (again both in and out) between Istanbul (TR1) and Western Black Sea (TR8), Western Marmara (TR2), Eastern Marmara (TR4) and Eastern Black Sea (TR9) regions. Except for Eastern Black Sea region (TR9), the regions of constant migration can be considered as the hinterland of Istanbul, thus the importance of proximity is also evident in these charts. The same persistent migration pattern and proximity effect can also be observed between Southeast Anatolia (TR C) and Mediterranean (TR 6), visualising the migration of predominantly Kurdish populations in the southeast to south. The migration flow from east to west, subject to many migration studies in Turkey, is also observable in the charts. Links from Northeast Anatolia (TR A), Central-east Anatolia (TR B) and Southeast Anatolia (TR C) to Istanbul (TR1) and to a certain extent to Aegean (TR 3) show this constant flow, even though harder to track due to low population density in the out-migration regions.

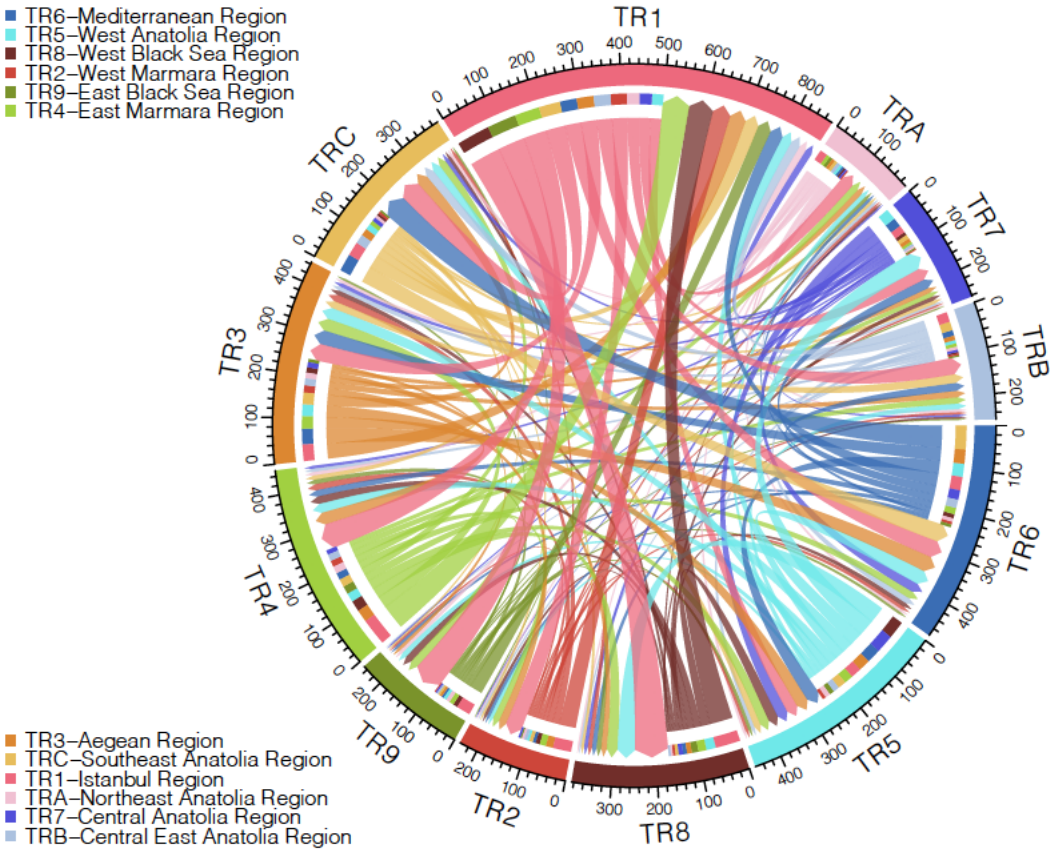

Figure 2.1: Bilateral migration flows at NUTS1 level in Turkey, 2019

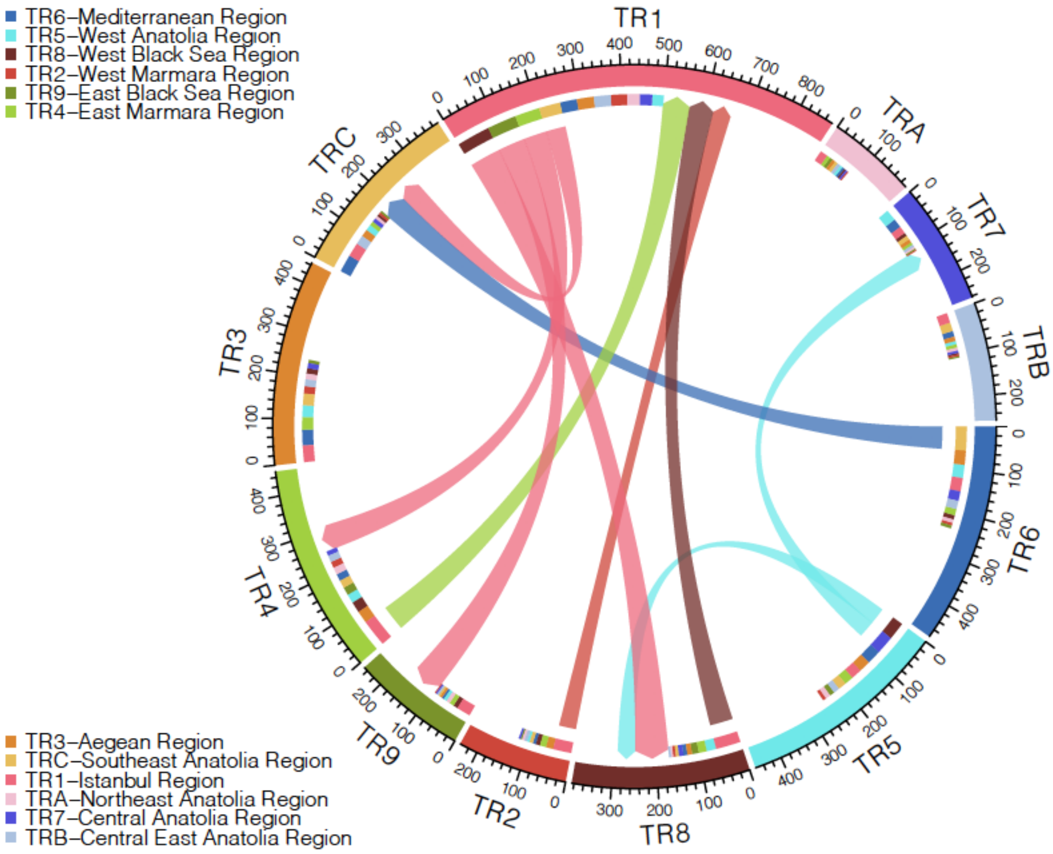

Figure 2.2: Top 10 largest bilateral migration flows at NUTS1 level in Turkey (thousands), 2008

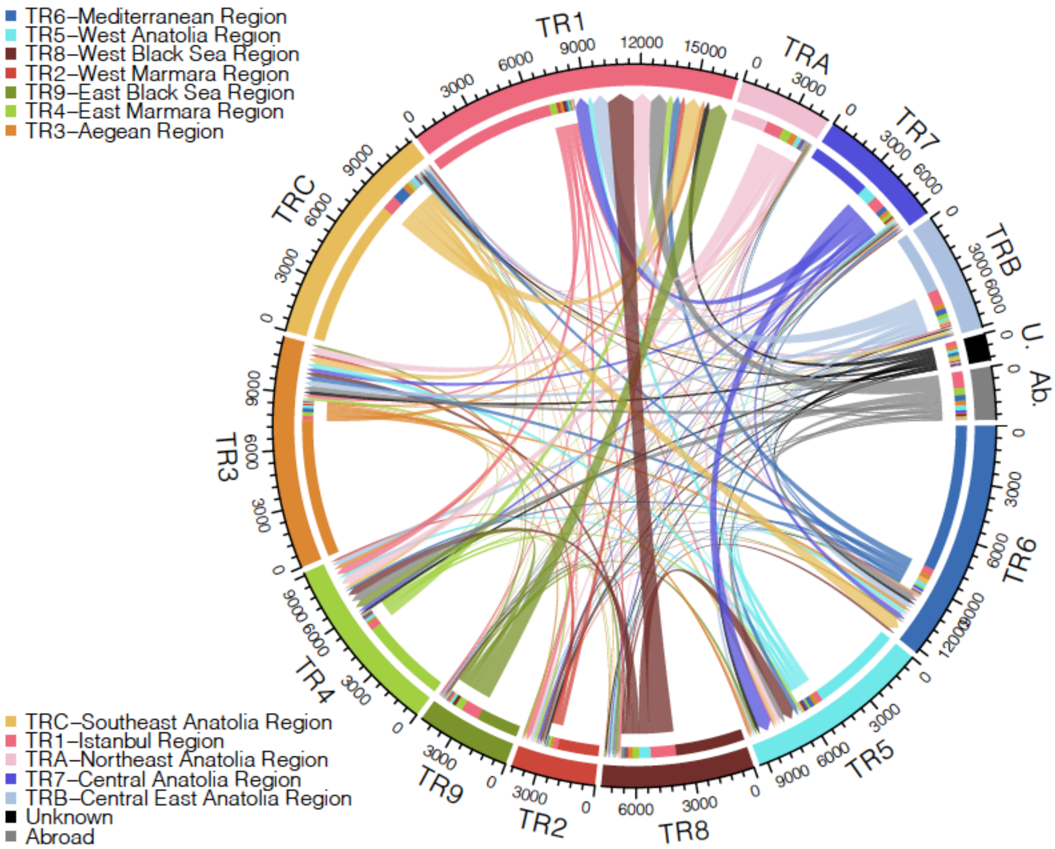

Figure 2.3: Origins of migration stocks at NUTS1 level in Turkey (thousands), 2019

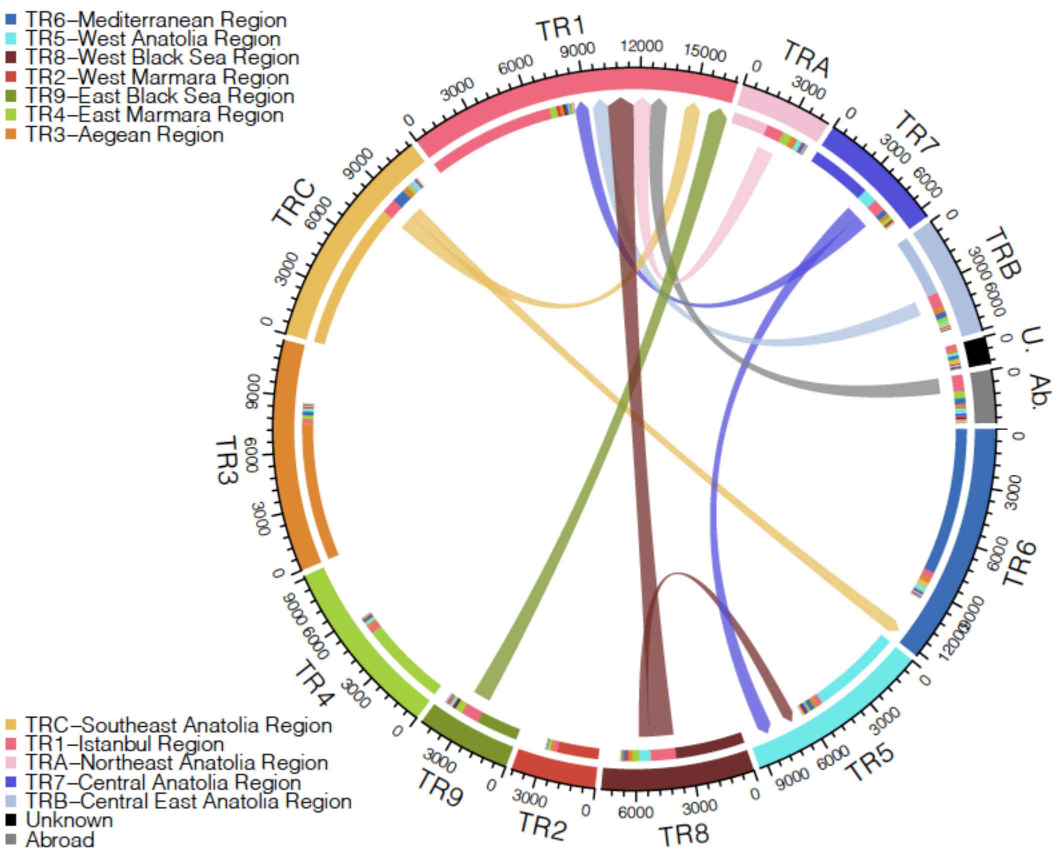

Figure 2.4: Top 10 Origins of migration stocks at NUTS1 level in Turkey (thousands), 2019

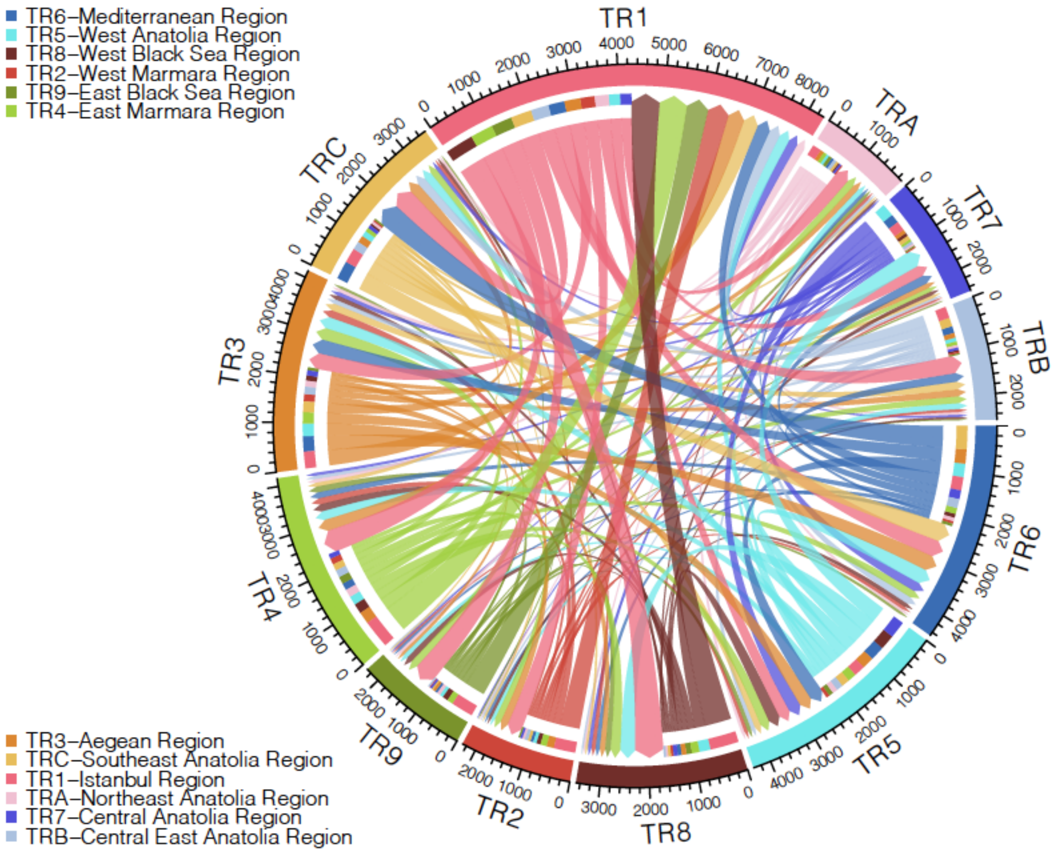

Figure 2.5: Cumulative bilateral migration flows at NUTS1 level in Turkey (thousands), 2010 - 2019

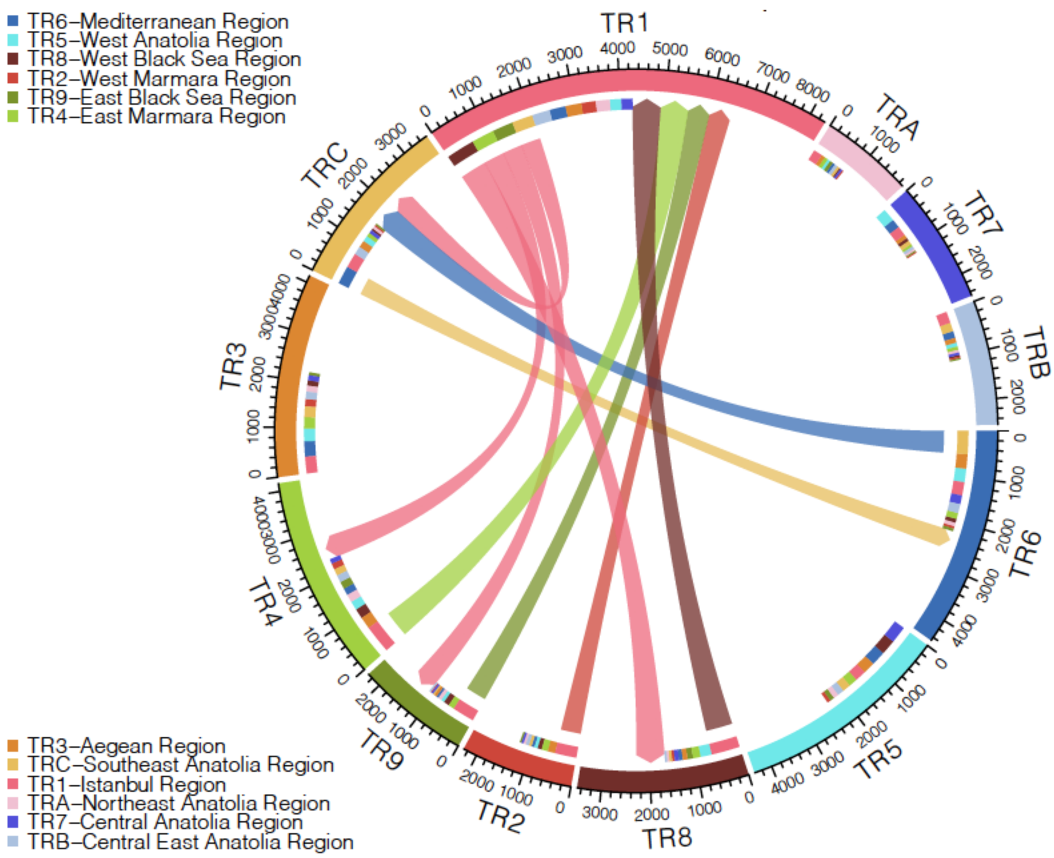

Figure 2.6: Top 10 largest bilateral migration flows at NUTS1 level in Turkey (thousands), cumulative for 2010 - 2019

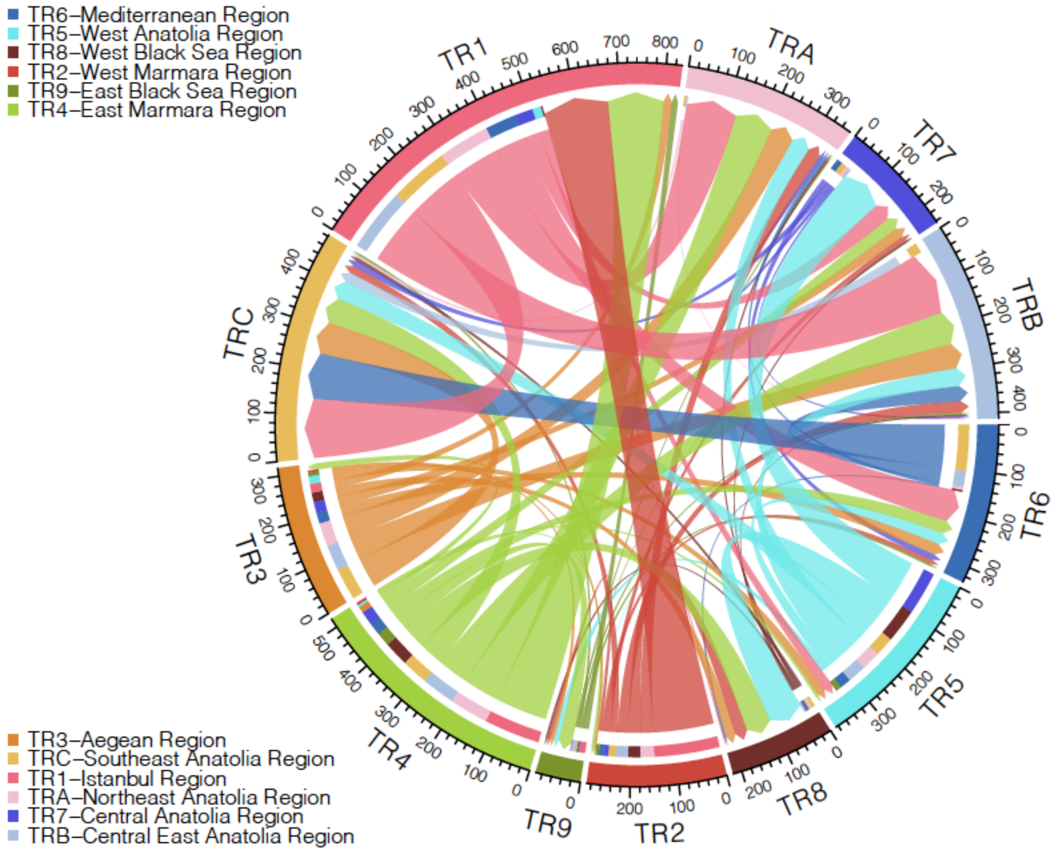

Figure 2.7: Cumulative net migration flows at NUTS1 level in Turkey (thousands), 2010 - 2019

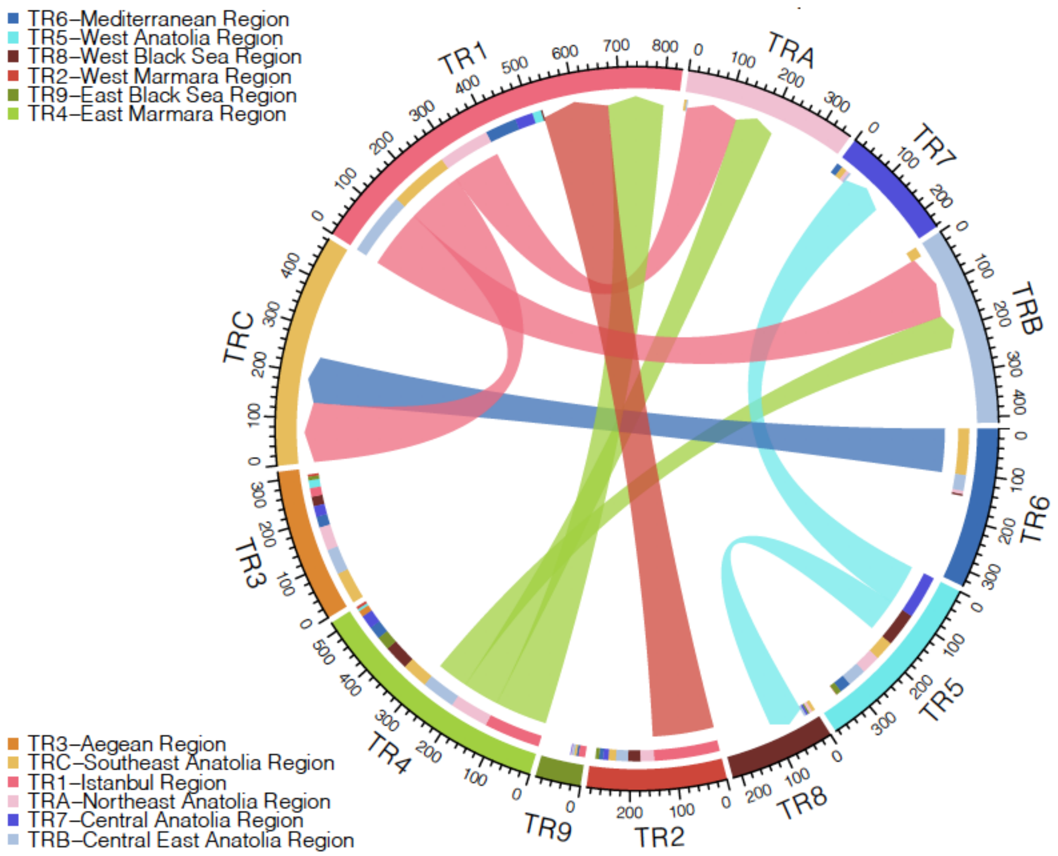

Figure 2.8: Top 10 net migration flows at NUTS1 level in Turkey (thousands), cumulative for 2010 - 2019

There are certain limitations to this analysis of internal migration patterns. First, due to the nature of the data we analysing the internal migration based on quantity, therefore, while it is an accurate method to track the relocation within the country, not knowing the purpose of migration limits the interpretation. Second, short-term migration such as seasonal agricultural workers are seemingly not represented as the short-term relocations are only shown in ABPRS data if the person registers his/her address at the neighbourhood[5] administration. This may also create an under-reporting of migration for educational purposes, especially the case of university students studying out of their hometowns. However, the frequency of elections in the period of analysis may create an advantage to mitigate the issue of under-reporting in this case. Considering that in Turkey, people are automatically registered as voters at the age of 18, but can vote only at the closest ballot station they are assigned to in their neighbourhood, and also considering that there were three general, two local elections and two referenda between 2008-2017, it is very likely that non-permanent relocations are also included in the data.

It should also be considered that in Turkey public servants can be relocated/reassigned to different provinces at will or by central decision, and are also required to complete a mandatory service in the disadvantaged regions for a period of two to three years. As they also register their new, but non-permanent addresses for official records, the role of public servants in the internal migration patterns shown above should also be considered. The case of public servants can also help to explain the out-migration from metropolitan areas/more developed regions, together with the possible movement of students who graduate and continue their education or start looking for a job in a different city or back in their hometowns.

Following the internal migration patterns, I proceed with the migration patterns of Syrians uTP in Turkey. As the data on the distribution of Syrians uTP across provinces starts at the end of 2015, the first diagram shows the settlement of refugees in NUTS-I regions in Turkey in 2015 (Figure 2.9 below). The following diagram,Figure 2.10, depicts the change in the stock of Syrians uTP by NUTS-1 regions, to better illustrate the change occurred in the four years. Similar to the approach taken for the Chord diagrams, further annual plots are presented in the Appendix 2.5, figures 2.35-2.38.

[5]Address registry in Turkey can be made at the neighbourhood administration (muhtarlık) level.

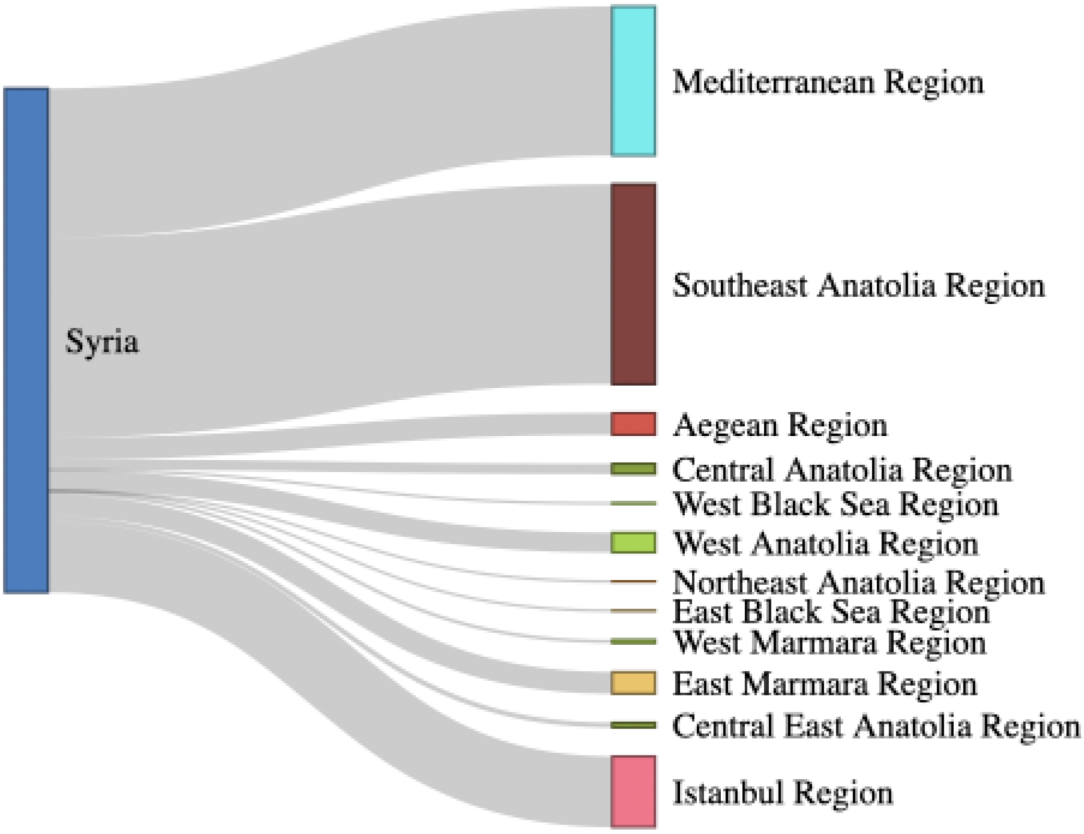

Figure 2.9: Distribution of SuTP at NUTS1 level, 2015

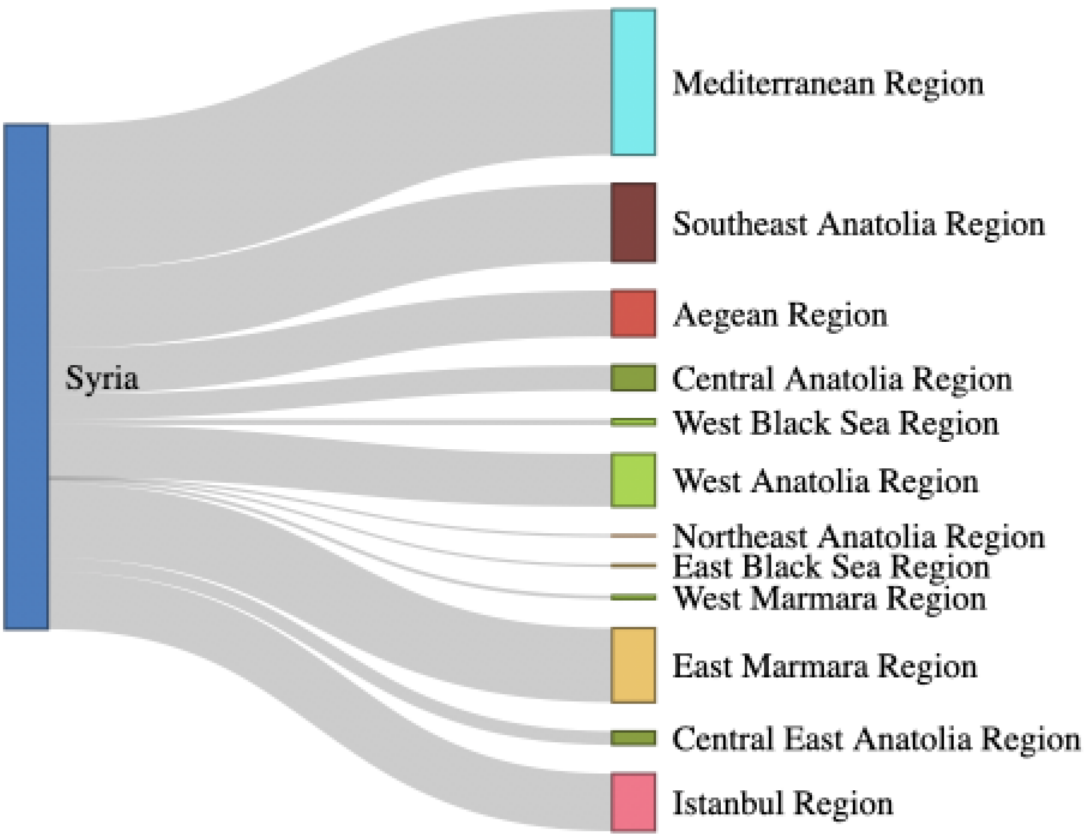

Figure 2.10: Increase in the # of SuTP at NUTS1 level, 2015-2019

The Sankey diagrams clearly show that Syrians under TP in Turkey are concentrated in three regions; Southeast Anatolia, Mediterranean and Istanbul. The distribution of Syrians under TP mapped over the years (Figures 2.39-2.43 in the Appendix 2.5) also confirms this pattern. The selection of the first region, Southeast Anatolia, is arguably due to proximity effect and cultural ties, since it is a region bordering Syria, where a minority of ethnic Arabs live. The second region Mediterranean is also under the proximity effect, as it borders Hatay, one of the entry points from Syria to Turkey. Population in Hatay is also very diverse, comprising of Turks and Kurds (Alawits and Sunnis) and also ethnic Arabs. The third prominent region is Istanbul, for which the proximity effect is not valid. Located far from the Syrian border, the pull effect of Istanbul for refugees is possibly the same for Turkish citizens, i.e. economic reasons.

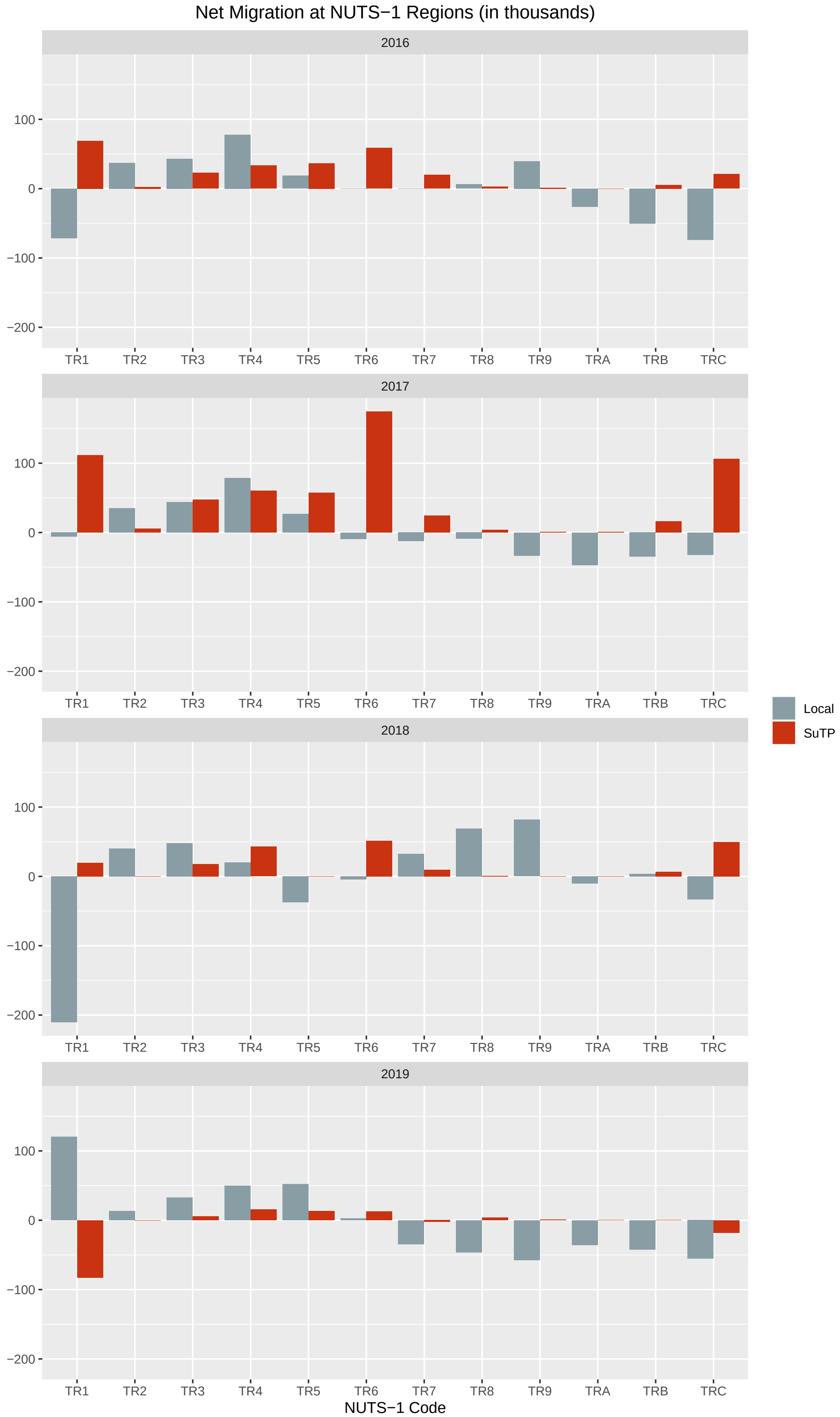

Figure 2.11: Comparison of locals and Syrians uTP in net migration at NUTS-1 regions (2016-2019)

The Figure 2.11 demonstrates a comparison between the absolute net migration in NUTS-I regions of Turkey in terms of both Turkish citizens (shown in gray) and Syrians uTP (shown in red). The bar plot shows that net migration patterns of Turkish citizens and Syrians uTP not necessarily follow the same trend or even the same direction. The Southeast Anatolia Region (TRC) is a known negative net migration region in Turkey, due to the Kurdish conflict, relatively younger population and relative underdevelopment. However, the region is bordering Syria and holds the strongest cultural ties with Syrians and therefore a more popular destination for Syrians uTP than Turkish citizens. Similarly, Mediterranean Region (TR6), while economically more developed than the Southeast Anatolia Region (TRC), is not as popular a destination region for Turkish citizens as it is for Syrians uTP. Comparing at the rankings of regions in terms of net migration each year (Faggian and Franklin, 2014), one can see that the top receiving and sending regions do not overlap for Turkish citizens and Syrians uTP.

The Figure 2.10 also illustrates that in addition to these areas, the Syrians uTP also started to move to East Marmara (TR4) region and West Anatolia (TR5) region between the years 2016-2019. As the assumptions of proximity to the border and cultural ties are not valid for these two regions, other factors such as economic seem to be affecting their settlement choices. These two regions are relatively more industrial than the Southeast Anatolia region at the Syrian border. It is also interesting that while both East Marmara (TR4) region and West Anatolia region (TR5) are almost always positive net migration regions 2.11, when examined in terms of bilateral flows, they are among the top negative (bilateral) net migration regions in the last decade, as demonstrated in Figure 2.8. These seemingly conflicting results for East Marmara (TR4) and West Anatolia (TR5) regions show that while they continue to attract both Turkish citizens and Syrians uTP, potentially due to the industrial job market, they maintain strong and dynamic mobility ties with specific regions.

2.3 Empirical Analysis

2.3.1 Data and Methods

Data used in the empirical analysis is obtained from three main sources. The data on the number of Syrians uTP across provinces over the five years between 2015-2019, which is the dependent variable of the statistical analysis, are obtained from DGMM registries. Almost all other indicators, constituting the explanatory and control variables in the analysis are gathered from Turkish Statistical Institute online statistics database. The only indicators that are not from the Turkish Statistical Institute database, are the minimum and maximum distance (in km) and driving duration (in hours) between provinces and the entry points at the Syrian border. The distance and duration measures are obtained via *googleway* package on R using the Google Maps Distance Matrix API.

Using the sources above, I created a panel data set combining demographic data and data on Syrians uTP at province level (81 provinces) for the years 2015-2019. In line with the requirements of the EU-Turkey Deal, general data on Syrian refugees is publicly available since early 2016 and updated weekly. The oldest available province-level data on Syrians uTP dates to January 15, 2016, which is considered as the data for the end of 2015. For all other years, annual data on Syrians uTP is taken as the last available data for each years. Although data on Syrians uTP are updated frequently, data taken from Turkish Statistical Institute is annual and in order to maintain consistency, end-of-the-year data is included in the panel.

The indicators obtained from the database of Turkish Statistical Institute and included in the panel data for analysis are; population, net migration rate, child dependency ratio, elderly dependency ratio, median age, Gini coefficient, GDP per capita[6], historical rate of minority languages, unemployment rate and employment rate. All variables except for the unemployment rate and employment rate are at province level. Data on unemployment rate and employment rate are only available at NUTS-2 level (26 sub-regions), therefore provinces located at the same sub-region are considered to have the same rates in the data set of this study. Although, this may create a certain degree of bias, considering that provinces in a sub-region have common socio-economic characteristics, and the number of regional-level indicators are limited, the bias is assumed to be negligible.

The distribution of immigrants and/or native speakers of immigrants' mother tongue is often used in the literature as a proxy for cultural ties and incentive for the settlement preferences of migrants. Historical settlement pattern of immigrants and/or historical rate of ethnic minorities are often used as an instrumental variable in studies aiming to assess the effect of migration on an issue of interest (Eugster et al., 2017). In the case of Turkey and Syrians uTP and election outcomes, for instance, historical distribution of Arabic-speaking minority in Turkey is used as an instrumental variable (Altındağ and Kaushal, 2020). As the aim of this study is to analyse the settlement preferences Syrians uTP in Turkey, the historical share of native Arabic speakers, measured in 1927, is used as an explanatory variable to account for the share of Syrians uTP across provinces.

In contrast to Altındağ and Kaushal (2020), I use the historical share of native Arabic speakers, as well as Kurdish and Circassian speakers for controls, across provinces based on 1927 census, instead of the 1965. The reason for this choice is that the 1927 census was the first census under the Republic, conducted in order to provide a clear picture of the demographics of the new state. In the following years, nation-state policies intensified, unity of language playing an important role. Thus, I believe that the first available province-level share of native Arabic speakers instead of the last available province-level share of native Arabic speakers would capture the historical share more accurately. However, I also included 1965 census data for robustness checks. In both cases, census data is obtained from TurkStat website.

One drawback of the 1927 and 1965 censuses is that it does not include information on all

[6]GDP data at province level is available only until the end of 2018

current provinces, as some have been recognized as a provincial administrative unit after the 1980s and 1990s. As the 1927 data is rigorous, providing information not only at province but also at district level, I started creating the variable of historical share of Arabic speakers at district level. This way, I re-assembled the current administrative units and managed to build a province-level variable in the most accurate way possible with the available data. The only missing province is Hatay, a border province with Syria, which became part of Turkey in 1939. Therefore, I used 1945 census for Hatay, to account for the earliest available historical share of native Arabic speakers. The map showing the distribution of the number of native Arabic speakers across provinces in 1927 (Figure 2.44) is provided in the Appendix (2.5). As the Figure 2.44 is based on actual numbers of native Arabic speakers, Hatay and Yalova[7] provinces are missing. In (Figure 2.45) the share of native Arabic speakers, relative to the province population is mapped, based on complete data used in the empirical analysis. Overall the share of Turkish citizens with Arabic as mother-tongue is 1.85% according to 1927 census and 1.51% according to 1965 census. Summary statistics of all variables are shown in the Table 2.3 in the Appendix 2.5.

The main motivation of the empirical analysis is to observe, whether replacement migration can have an opportunity, thus whether it could be possible to incentivise policies for Syrians uTP to settle in provinces where the dependency ratio is high. By the descriptive analysis using data visualisation methods, we observe two particular trends in the settlement preferences of Syrians uTP. The first trend is to settle in provinces close to the Syrian border, as a combination of the proximity effect (Ravenstein, 1885) and cultural ties (Eugster et al., 2017). The second trend, which is also observed to become intensified in recent years, is to settle in provinces with industrial development and/or relatively more dynamic economic activity. This trend appears closer to the economic motivations in internal migration patterns. Considering the EU-Turkey Deal that came into effect in March 2016 and closed-off the passageway for Syrian refugees from Turkey to the EU and obliging them to stay in Turkey under the status of temporary protection, it is possible that after this date and gradually the settlement patterns of Syrians uTP came to be shaped more by economic motivations and approximate internal migration patterns. Thus, the hypothesis tested here is as follows;

H1: The number of Syrians uTP in a province is positively correlated with the migration behaviour of Turkish citizens for that province

To analyse the factors that affect the settlement pattern of Syrians uTP, I use a panel data approach. I regress the independent variables of natural logarithm of population, net migration rate, median age, historical share of Arabic-speaking minority, distance from the Syrian border, child dependency ratio, elderly dependency ratio, unemployment rate, employment rate on the

[7]In 1927 central Yalova used to be part of a district (Karamürsel) in Kocaeli. For the share of native Arabic speakers indicator, the share in Kocaeli is considered as no administrative unit of Yalova as of today is included in the 1927 census data

dependent variable of the natural logarithm of the number of Syrians uTP in a province. The main aim is to observe the socio-economic factors that contribute to the settlement choices of Syrians uTP. In this quest, I use the population and net migration rate of provinces as indicators of internal migration patterns. The historical share of Arabic-speaking minority is considered as a proxy of cultural ties and distance from the Syrian border measures the proximity effect. Unemployment and employment rate are considered as indicators of the economic conditions in a province. Last, the variables of median age, child dependency ratio, elderly dependency ratio are regarded as indicators to help examining the possibility of replacement migration.

I apply both fixed effects and random effects models, however, the design of this study prefers random effects model, considering that there is a significant amount of variance between provinces and two time-invariant variables, historical rate of native Arabic speakers and distance to border entry points, are important variables of interest. In the fixed-effects model all province-specific time-invariant characteristics are absorbed, to enable observation of time-variant characteristics. Therefore, the fixed effects model is considered as a robustness check, in which the historical settlement pattern of Arabic-speaking minority and distance to border measures are omitted.

2.3.2 Results

Corroborating with the design of this study, the random effects model is preferred by the Hausman test over the fixed effects model. The results of the random effects model are presented in Table 2.1, while the comparison with the fixed effects model is to be found in the Appendix (2.5).

The results shown in Table 2.2 reject the *H1* and show that the association between the number of Syrians uTP and the net migration rate of a province is negative and significant. Furthermore, the association between the number of Syrians uTP and the median age in a province is positively correlated, but loses its significance when the year fixed-effects are introduced. The negative association between the number of Syrians uTP and the net migration rate remains almost unchanged with the introduction of year fixed-effects.

Among the control variables, the (logarithmic transformation of) province population and the historical rate of native Arabic-speakers (1927) are found to have a positive and significant association with the number of Syrians uTP. The minimum distance to the border is significantly and negatively correlated with the number of Syrians uTP, indicating that geographic proximity to the border, similar to the cultural proximity shown by the historical rate of native Arabic-speakers (1927), is an incentive for settlement for Syrians uTP. Although the economic activity is concentrated in the west of the country and far from the Syrian border, the results imply that economic factors are not the only motivation in the settlement decisions of the Syrians uTP in Turkey. The association observed between the number of Syrians uTP and employment

Table 2.1: Results of the empirical analysis (Random effects model)

VARIABLES	(1) ln(Refugees)	(2) ln(Refugees)
Net Migration Rate	-0.00197**	-0.00195**
	(0.000950)	(0.000814)
Median Age	0.426***	0.119
	(0.129)	(0.146)
ln(Population)	1.791***	1.800***
	(0.221)	(0.212)
1927 Rate of Arabic Speakers	0.0205	0.0308**
	(0.0201)	(0.0148)
Unemployment Rate	0.0229**	0.00271
	(0.0109)	(0.0130)
Employment Rate	-0.00572	-0.0179*
	(0.00669)	(0.0104)
sqrt(Min. Distance)	-0.214***	-0.117***
	(0.0323)	(0.0383)
Child Dependency Ratio	0.0369	0.00904
	(0.0387)	(0.0371)
Elderly Dependency Ratio	-0.173*	-0.0667
	(0.0894)	(0.101)
Gini coefficient	-1.509	-1.595
	(1.766)	(1.748)
Constant	-22.30***	-15.05***
	(3.515)	(3.689)
Observations	405	405
Number of provinces	81	81
Year FE	NO	YES

Robust standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

& unemployment rates across provinces also seem to support this argument. However, due to possible reverse causality bias, I refrain from reading more into the negative association between the number of Syrians uTP and employment rate.

To control the robustness of these results I replicated the random effects model with alternative variables, such as including historical rate (1927) of native Kurdish and Circassian speakers, replacing 1927 data for native Arabic speakers with 1965 census instead, including maximum distance and minimum/maximum driving duration to border instead of minimum distance. The results of these robustness checks are reported in the Appendix 2.5 under Table 2.4. An interesting result that can be observed by the alternative native language variables is that the correlation with the 1927 rate of native Kurdish speakers is significant and negative. Although the predominantly Kurdish provinces are located close to the Syrian border and the cultural proximity could be assumed, in comparison to for instance, Western predominantly Turkish provinces, the results suggest the opposite. Along with the difference in language, economic reasons and sectarian difference (for the provinces where Kurdish Alawites have a higher share in the population) may play a role in these results, as a possible area of future research. Last but not least, the rate of native Arabic speakers from 1965 census is observed as insignificant, signalling that using 1927 data has more potential to measure the cultural ties.

Furthermore, I replicated the model after dropping Istanbul from the dataset, considering that Istanbul may confound the estimations (2.5). In all models, results of which are provided in the tables 2.4 and 2.5, the significance and direction of the main variables of interest are maintained.

2.4 Conclusion

The characteristics of internal and international migration patterns are often investigated separate from each other. However, the recent sizable influx of Syrian refugees, their long-term stay in urban areas and their need to find means of livelihood suggests a potential area of research, to investigate in which ways internal migration and refugee settlement patterns converge and diverge. In this study, I aimed to shed light to whether there is a correlation between the migration behaviour of Turkish citizens and settlement pattern of Syrians uTP. I consider the Syrians uTP in Turkey as a special case, one that bears the characteristics of both traditional refugee-policy approach and labour migration. Syrians uTP are supported by numerous international organizations and access to public services is enabled for free based on the EU-Turkey Deal and support of UNHCR. Yet, the share Syrians uTP in camps is quite low, and despite the very basic safety net provided by the efforts of international organizations and Turkish authorities, a great majority of them have to sustain their lives on their own. Under the status of temporary protection, they are free to move between provinces and are in need of jobs, a characteristic they share with job-seeking immigrants. Furthermore, they will remain in Turkey for medium- to long-term, and a certain share of Syrians uTP would probably never return to Syria ever.

These features of the Syrians uTP issue require immediate attention of both researchers and policy-makers to develop long-term and sustainable policies that would support Syrians uTP and would suit for the natives.

I believe evidence-based policies is important in every field but especially so, in cases like this. The main idea behind this study was to see if replacement migration could offer any solution to this issue. To see that, I questioned if the internal migration patterns and refugee settlement patterns match anyhow and tested this using data visualisation methods and a simple empirical analysis.

Data visualisation analysis demonstrated that a majority of Syrians uTP are concentrated in three regions; southeast Anatolia, Mediterranean and Istanbul. Compared to the internal migration patterns in Turkey, these results show that although the pull effect of Istanbul (also Izmir) is similar for Turkish citizens and Syrian refugees, apart from the metropolitan areas, the patterns show divergences. A concentration in southeast Anatolia and Mediterranean signals that proximity both in terms of geography and culture/language play a more important role for the determinants of settlement. In line with this, it should also be considered that within the Mediterranean region, refugee concentration is more towards east (Syrian border) rather than towards west, where tourism-driven economies (Antalya and Muğla) are located.

In the empirical analysis part, the main hypothesis, assuming a positive and significant correlation between net migration rate of provinces and number of Syrians uTP, is disproved. The association is found as negative and significant instead, suggesting that the migration behaviour of the locals and the Syrians uTP are in opposition. While these results provide a base for replacement migration policies to be developed, further research is needed to understand the characteristics of this negative association. Furthermore, the lack of proximity effect to border signals that economic activity is an important factor for the settlement patterns of Syrians uTP, similar to the locals.

The results of both data visualisation and empirical analysis imply that Syrian refugees prefer to settle in provinces that are culturally close to their own country and/or in provinces where they know job opportunities would be higher. However, this pattern still leaves room for replacement migration in depopulated areas. Considering the algorithm suggested by Bansak et al. (2018), the findings show that there is room for policy-making that would incentivize and relocate Syrian refugees where they would have a better chance at finding a job as well as integrate with the native inhabitants. Future research focusing on this issue and using alternative and more detailed sets of control variables, hence may have a chance to eventually develop a similar algorithm for the case of Turkey.

2.5 Appendix-Chapter 2

Table 2.2: NUTS (Nomenclature of Territorial Units for Statistics) classification in Turkey

Code	NUTS I – Regions	NUTS II – Sub-regions	NUTS III - Provinces
TR1	Istanbul	Istanbul	Istanbul
TR 2	Western Marmara	Tekirdağ, Balıkesir sub-regions	Tekirdağ, Edirne, Kırlareli / Balıkesir, Çanakkale
TR 3	Aegean	Izmir, Aydın, Manisa sub-regions	Izmir / Aydın, Denizli, Muğla / Manisa, Afyon, Kütahya, Uşak
TR 4	Eastern Marmara	Bursa, Kocaeli sub-regions	Bursa, Eskişehir, Bilecik / Kocaeli, Sakarya, Düzce, Bolu, Yalova
TR 5	Western Anatolia	Ankara, Konya sub-regions	Ankara / Konya, Karaman
TR 6	Mediterranean	Antalya, Adana, Hatay sub-regions	Antalya, Isparta, Burdur / Adana, Mersin / Hatay, Kahramanaraş, Osmaniye
TR 7	Central Anatolia	Kırıkkale, Kayseri sub-regions	Kırıkkale, Aksaray, Niğde, Nevşehir, Kırşehir / Kayseri, Sivas, Yozgat
TR 8	Western Black Sea	Zonguldak, Kastamonu, Samsun sub-regions	Zonguldak, Karabük, Bartın/ Kastamonu, Çankırı, Sinop / Samsun, Tokat, Çorum, Amasya
TR 9	Eastern Black Sea	Trabzon sub-region	Trabzon, Ordu, Giresun, Rize, Artvin, Gümüşhane
TR A	Northeast Anatolia	Erzurum, Ağrı sub-regions	Erzurum, Erzincan, Bayburt / Ağrı, Kars, Iğdır, Ardahan
TR B	Central-east Anatolia	Malatya, Van sub-regions	Malatya, Elazığ, Bingöl, Tunceli / Van, Muş, Bitlis, Hakkari
TR C	Southeast Anatolia	Gaziantep, Şanlıurfa, Mardin sub-regions	Gaziantep, Adıyaman, Kilis / Şanlıurfa, Diyarbakır / Mardin, Batman, Şırnak, Siirt

Figure 2.12: NUTS-1 level regions in Turkey

Table 2.3: Descriptive statistics

Variable		Mean	Std. Dev.	Min	Max	Observations
	overall	32.047	5.432	19.34	40.76	N = 405
Median Age	between		5.43	19.678	39.922	n = 81
	within		0.563	30.011	34.141	T = 5
Net Migration	overall	-0.902	17.908	-118.97	138.51	N = 405
Rate	between		7.699	-25.802	18.286	n = 81
	within		16.186	-122.116	133.260	T = 5
Population	overall	998,818.1	1803942	78550	1.55e+07	N = 405
	between		1,812,442	83,247.6	1.50e+07	n = 81
	within		42,274.26	640,697.7	1,502,531	T = 5
Unemployment	overall	10.752	5.486	3.8	31.1	N = 405
Rate	between		5.020	6.14	27.38	n = 81
	within		2.266	4.012	21.112	T = 5
Employment	overall	52.242	7.111	30	66.3	N = 405
Rate	between		6.391	34.04	61.58	n = 81
	within		3.180	45.742	64.602	T = 5
Dependency	overall	50.050	8.172	37.12	78.91	N = 405
Ratio	between		8.155	38.572	77.776	n = 81
	within		0.967	46.376	54.336	T = 5
Child	overall	35.204	11.549	19.6	72.29	N = 405
Dependency	between		11.562	20.026	71.05	n = 81
Ratio	within		1.006	31.426	39.456	T = 5
Elderly	overall	14.846	4.895	4.77	29.16	N = 405
Dependency	between		4.890	4.934	28.348	n = 81
Ratio	within		0.537	13.4	16.608	T = 5

Table 2.3 cont.

Variable		Mean	Std. Dev.	Min	Max	Observations
Gini	overall	0.361	0.0226	0.309	0.43	N = 405
Coefficient	between		0.018	0.331	0.41	n = 81
	within		0.014	0.327	0.399	T = 5
SuTP	overall	39,441.66	97,067.02	14	557,663	N = 405
	between		96,438.61	35.2	470,731.6	n = 81
	within		14,617.97	-74,064.94	126,373.1	T = 5
1927 Share of	overall	1.852	6.626	0	39.218	N = 405
Native Arabic	between		6.659	0	39.218	n = 81
Speakers	within		0	1.852	1.852	T = 5
1965 Share of	overall	1.511	5.01	0	29.254	N = 405
Native Arabic	between		5.035	0	29.254	n = 81
Speakers	within		0	1.511	1.511	T = 5
1927 Share of	overall	15.596	26.642	0	84.546	N = 405
Native Kurdish	between		26.775	0	84.546	n = 81
Speakers	within		0	15.595	15.595	T = 5
1927 Share of	overall	0.778	2.004	0	15.451	N = 405
Native Circ.	between		2.014	0	15.451	n = 81
Speakers	within		0	0.778	0.778	T = 5
Minimum	overall	644.837	348.373	9.77	1.379.25	N = 405
Distance to	between		350.111	9.77	1.379.25	n = 81
Border (km)	within		0	644.836	644.836	T = 5
Maximum	overall	979.729	355.471	331.74	1,744.58	N = 405
Distance to	between		357.243	331.74	1,744.58	n = 81
Border (km)	within		0	979.729	979.729	T = 5
Minimum	overall	7.713	3.687	0.261	14.528	N = 405
Duration to	between		3.705	0.261	14.528	n = 81
Border (hrs)	within		0	7.713	7.713	T = 5
Maximum	overall	11.108	3.572	4.105	17.982	N = 405
Duration to	between		3.589	4.105	17.982	n = 81
Border (hrs)	within		0	11.108	11.108	T = 5

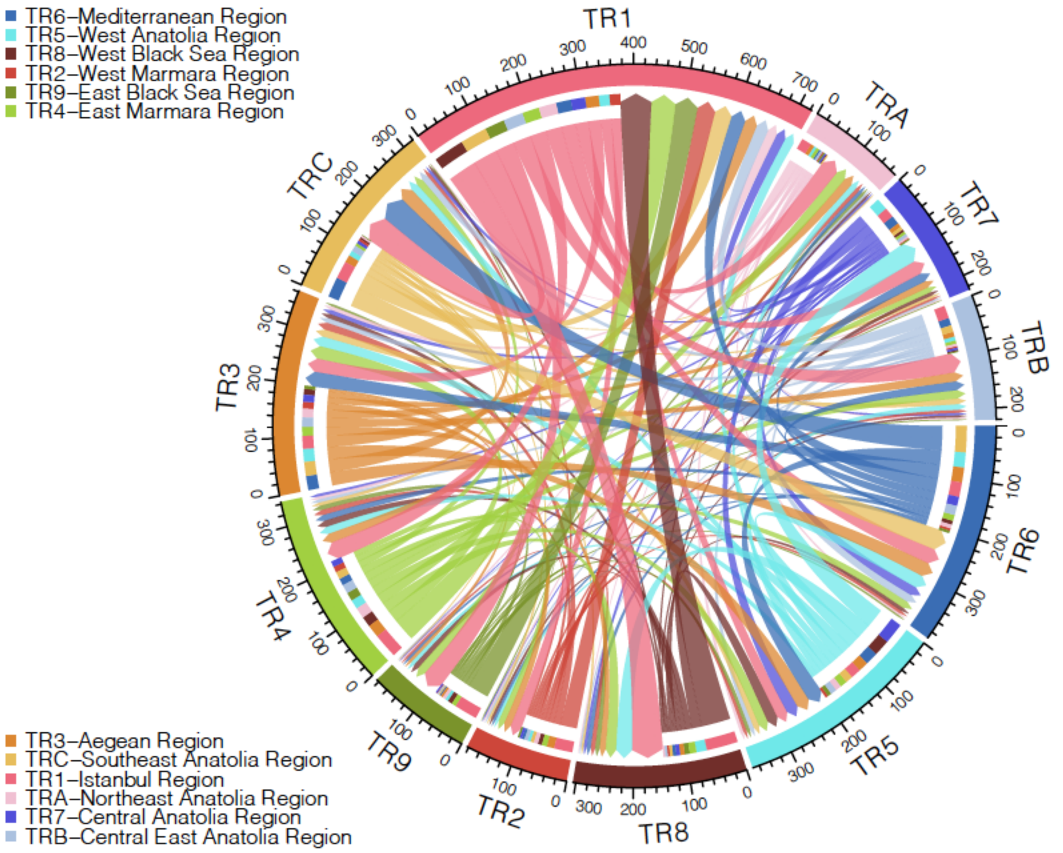

Figure 2.13: Bilateral migration flows at NUTS1 level in Turkey (in thousands), 2008

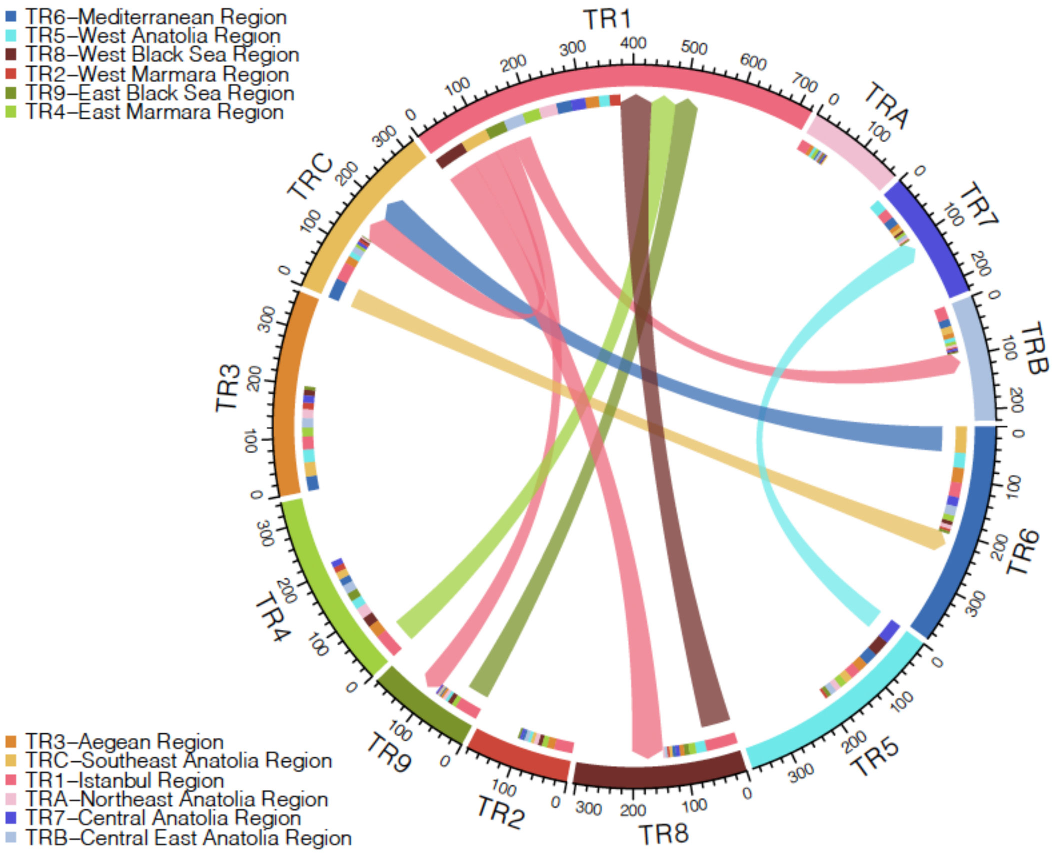

Figure 2.14: Top 10 largest bilateral migration flows at NUTS1 level in Turkey (in thousands), 2008

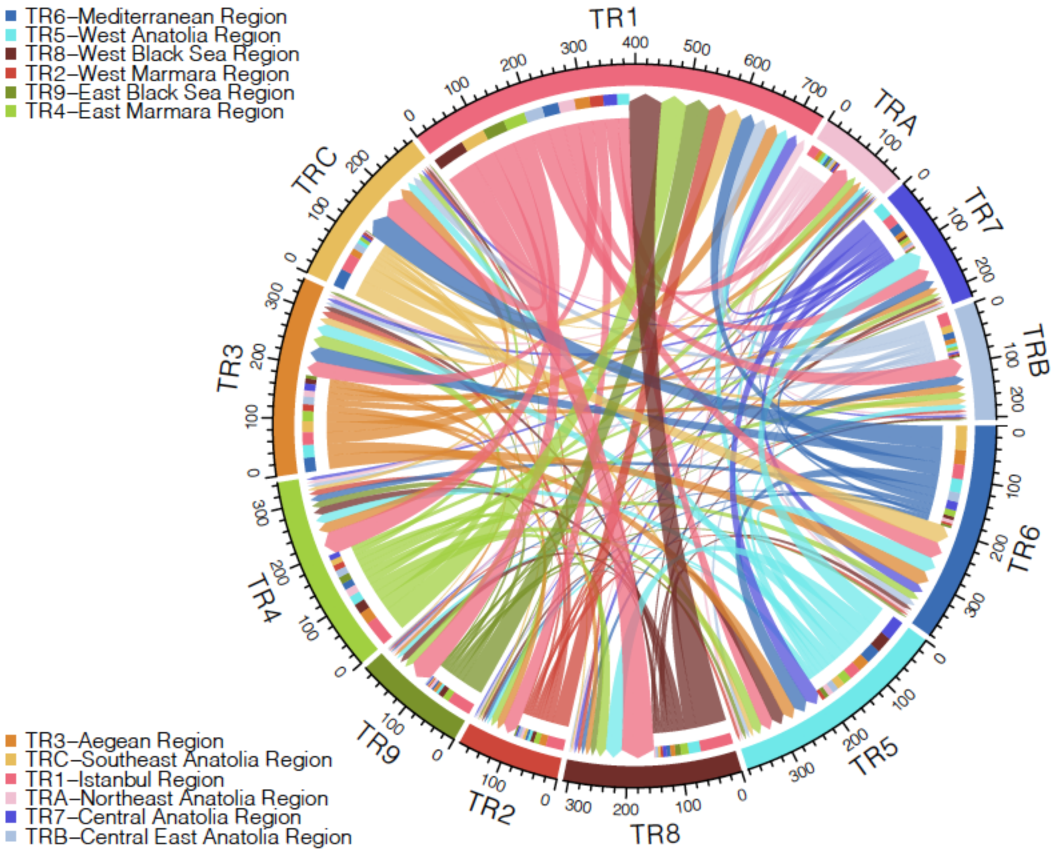

Figure 2.15: Bilateral migration flows at NUTS1 level in Turkey (thousands), 2009

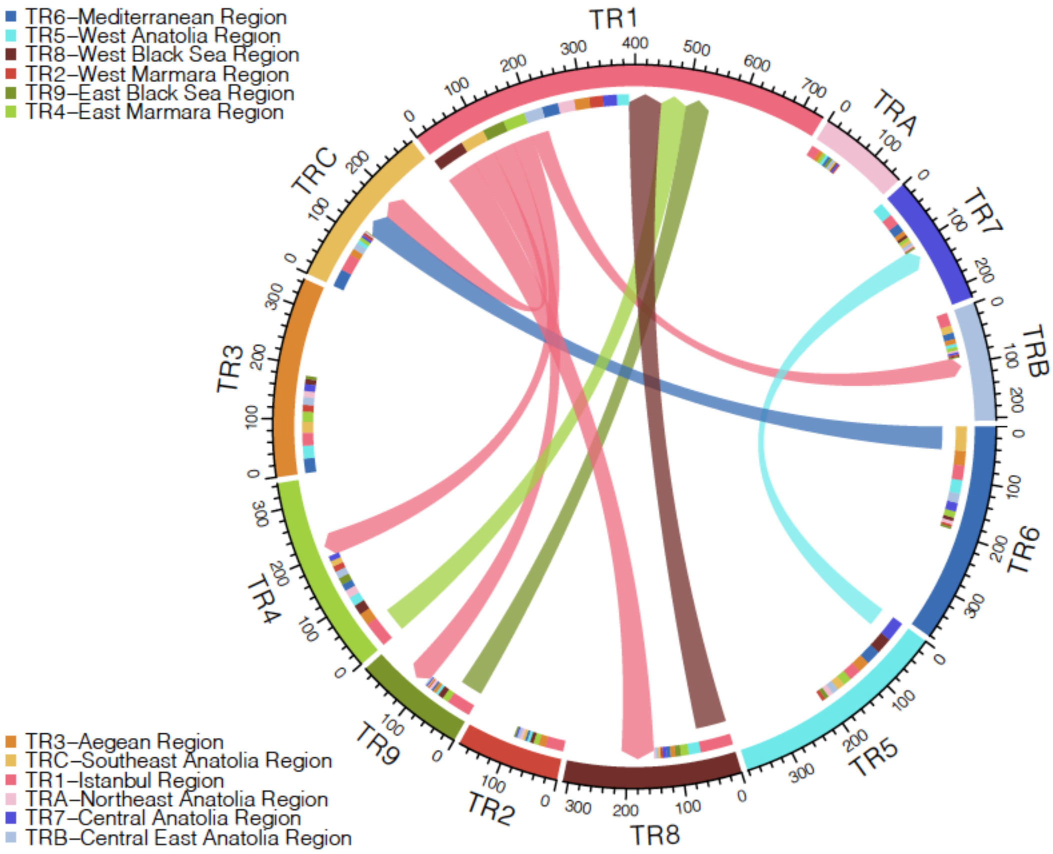

Figure 2.16: Top 10 largest bilateral migration flows at NUTS1 level in Turkey (thousands), 2009

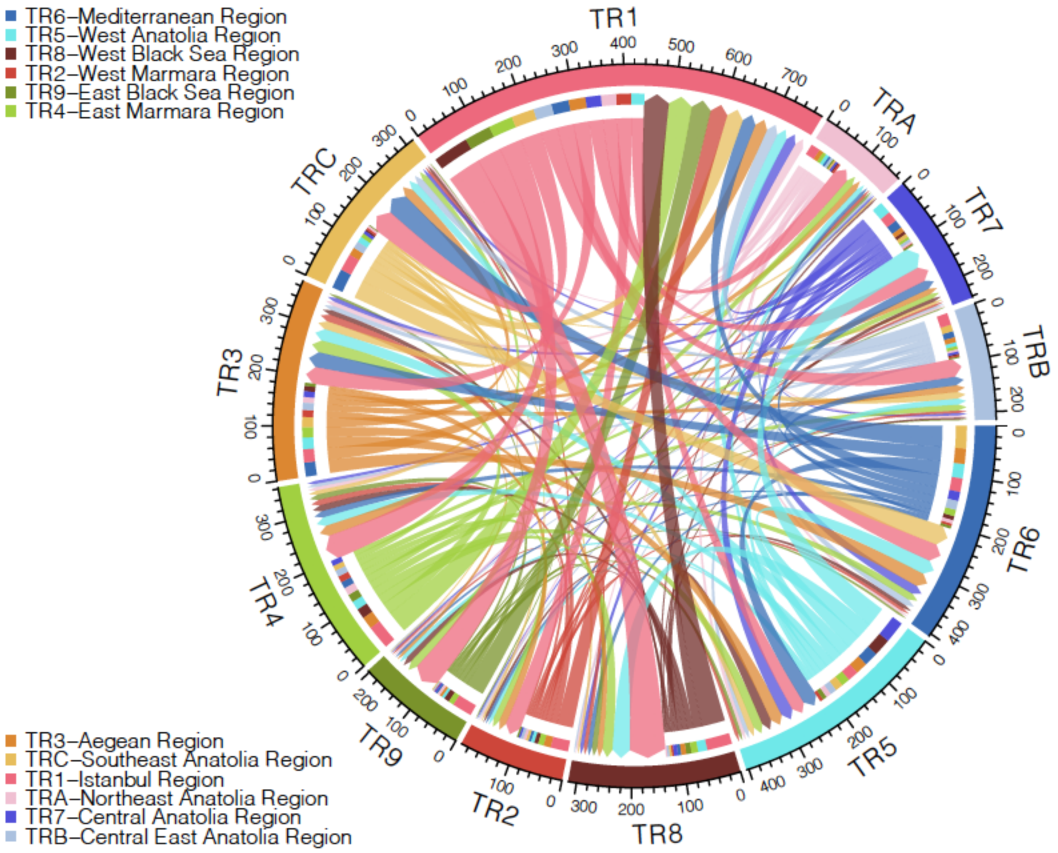

Figure 2.17: Bilateral migration flows at NUTS1 level in Turkey (thousands), 2010

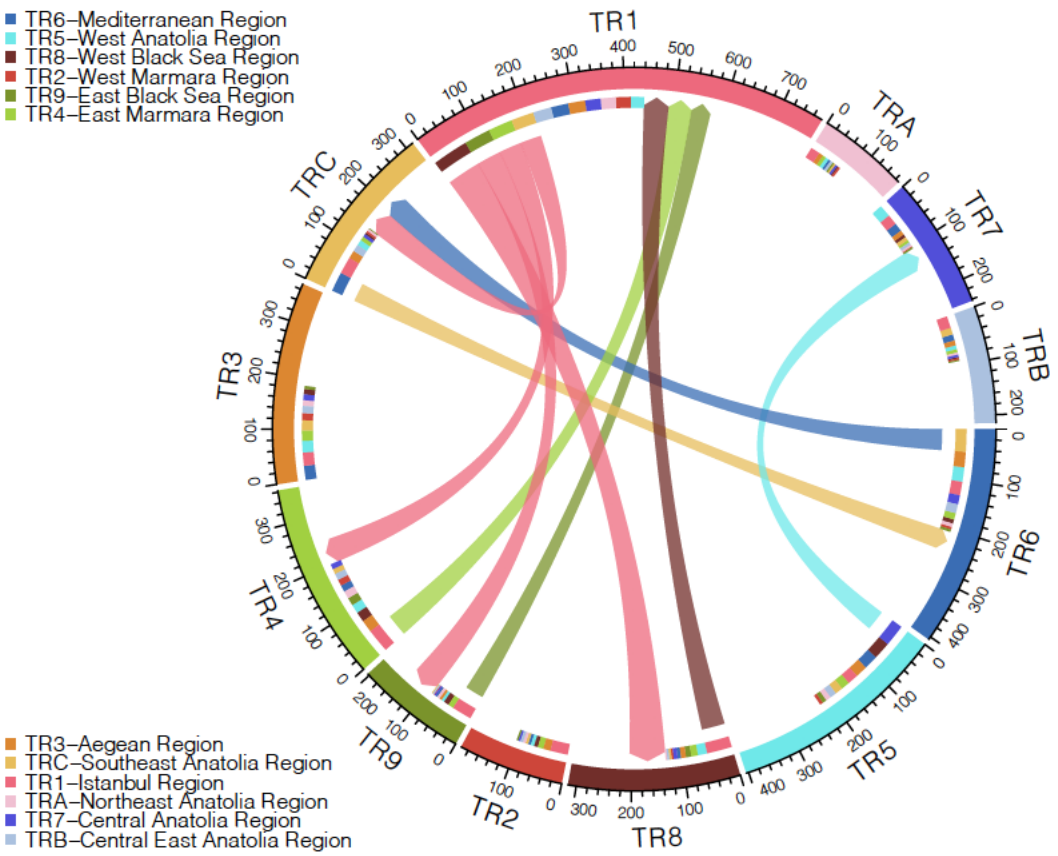

Figure 2.18: Top 10 largest bilateral migration flows at NUTS1 level in Turkey (thousands), 2010

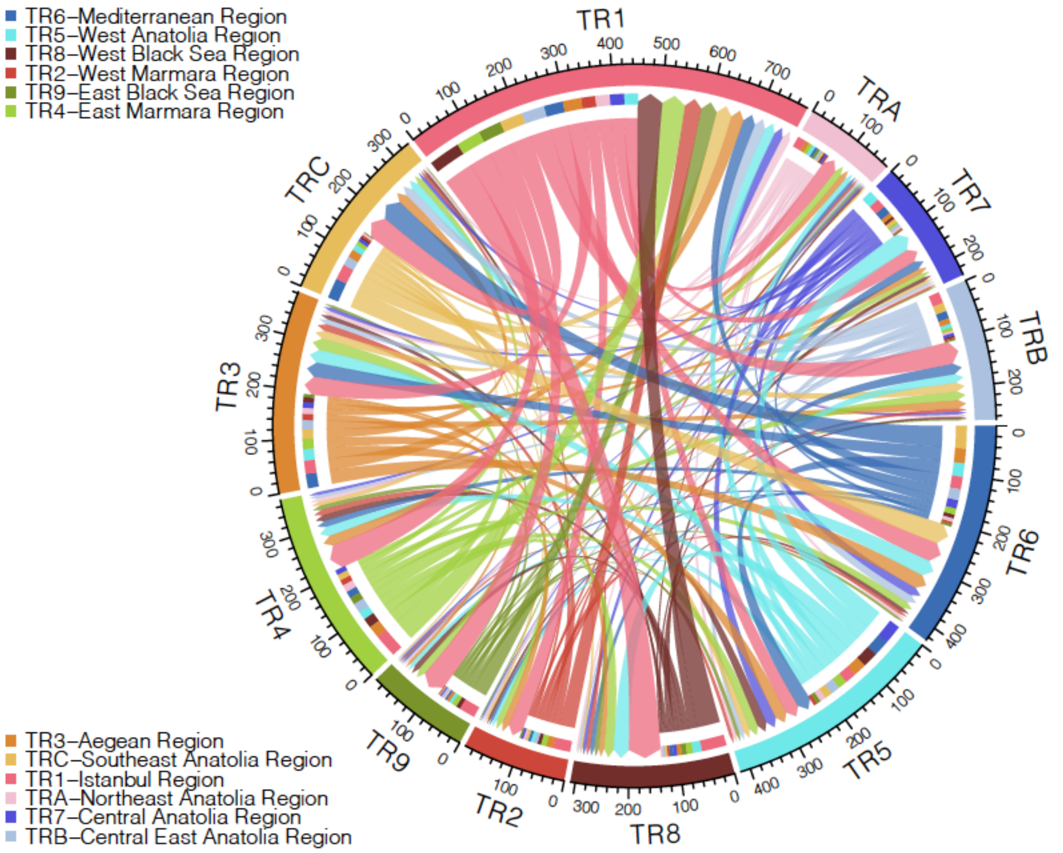

Figure 2.19: Bilateral migration flows at NUTS1 level in Turkey (thousands), 2011

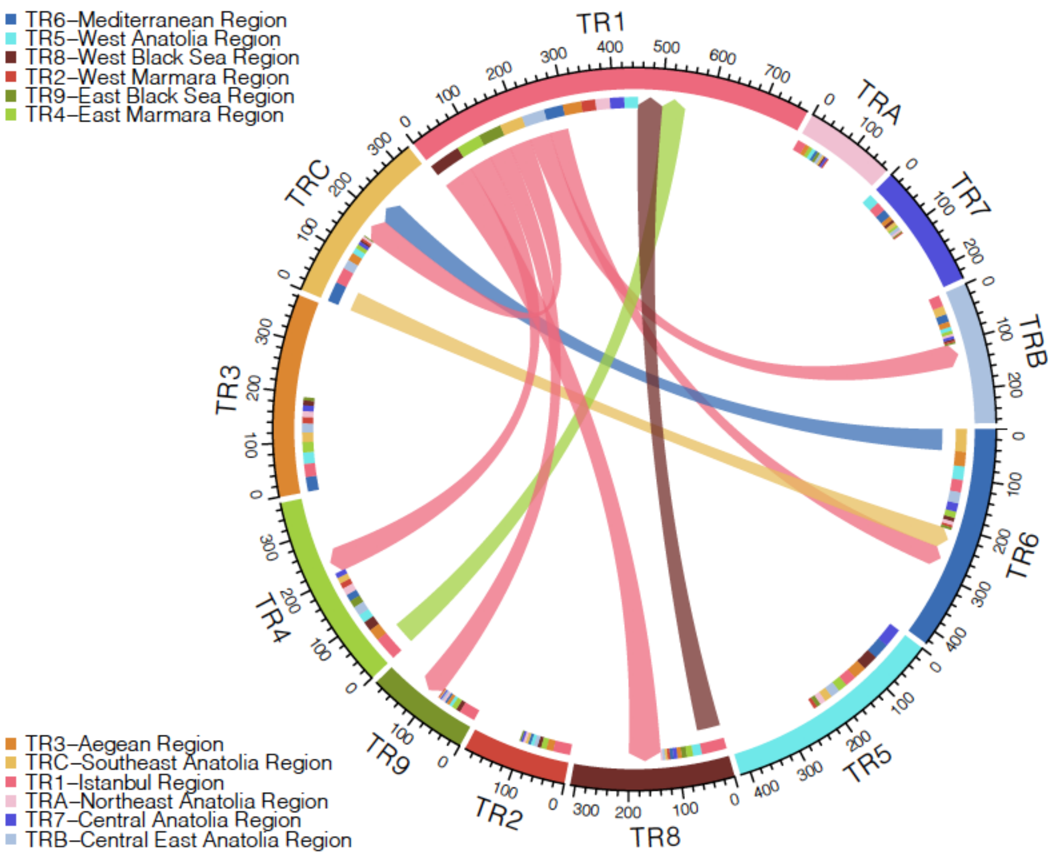

Figure 2.20: Top 10 largest bilateral migration flows at NUTS1 level in Turkey (thousands), 2011

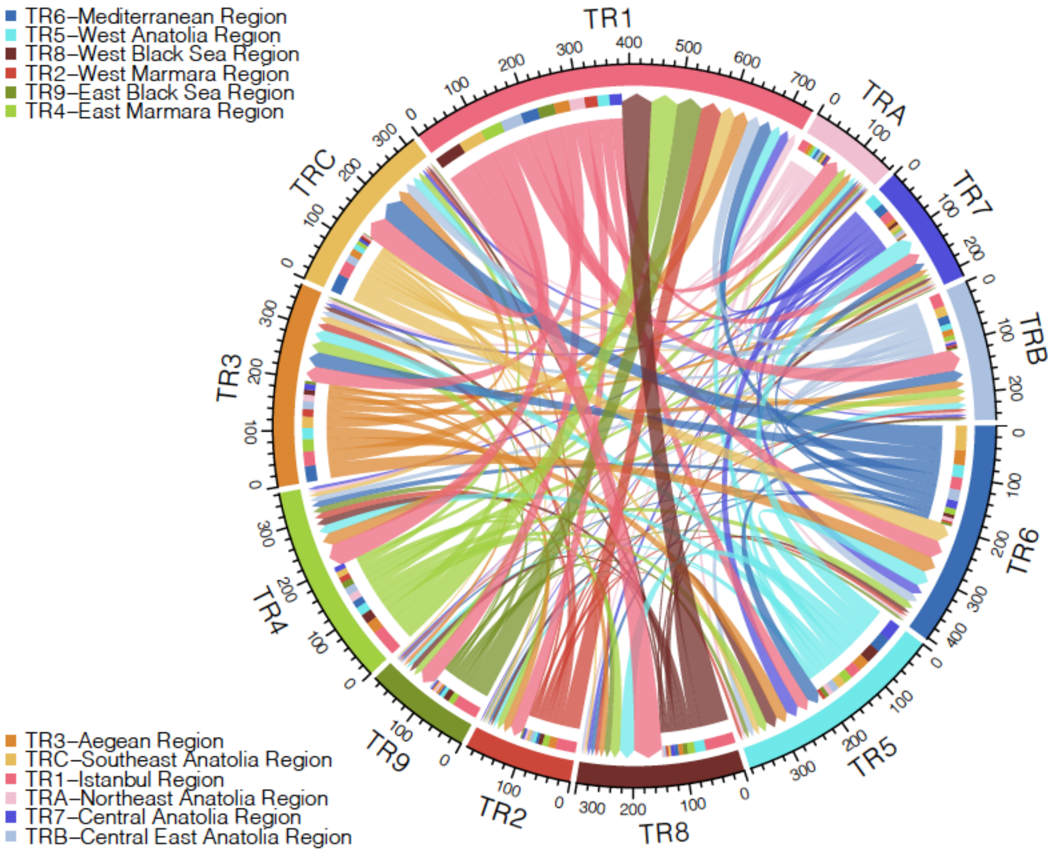

Figure 2.21: Bilateral migration flows at NUTS1 level in Turkey (thousands), 2012

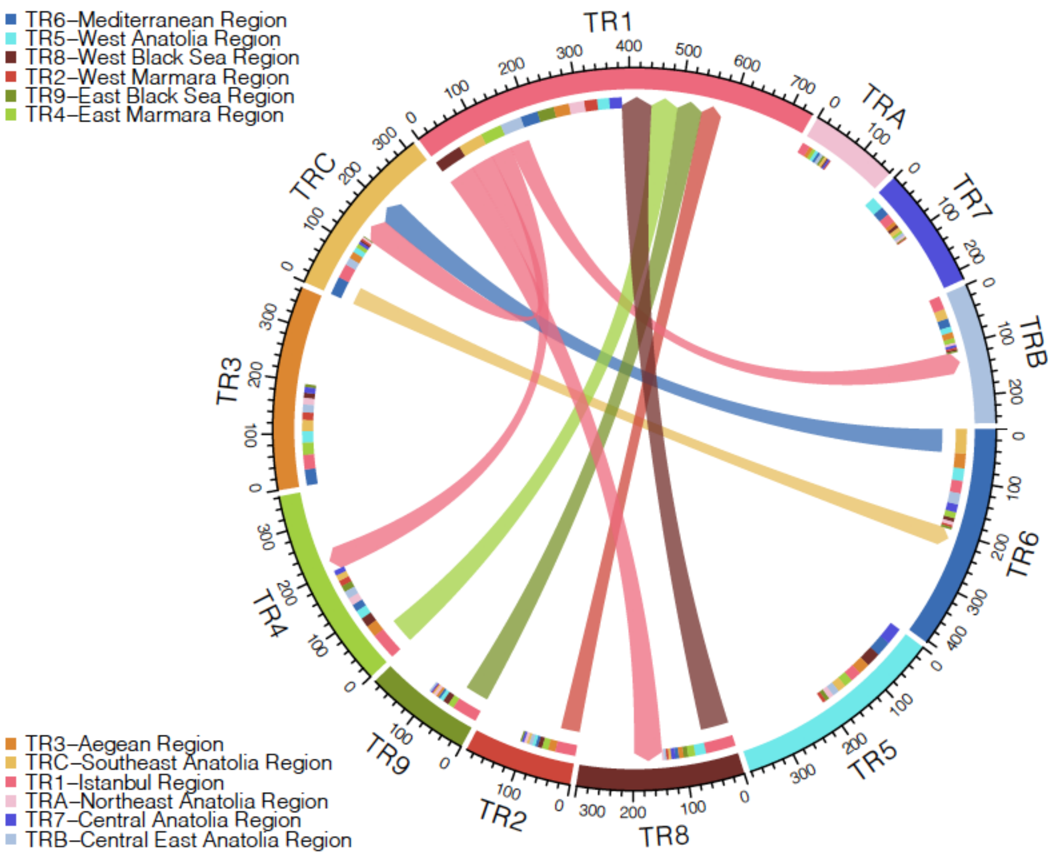

Figure 2.22: Top 10 largest bilateral migration flows at NUTS1 level in Turkey (thousands), 2012

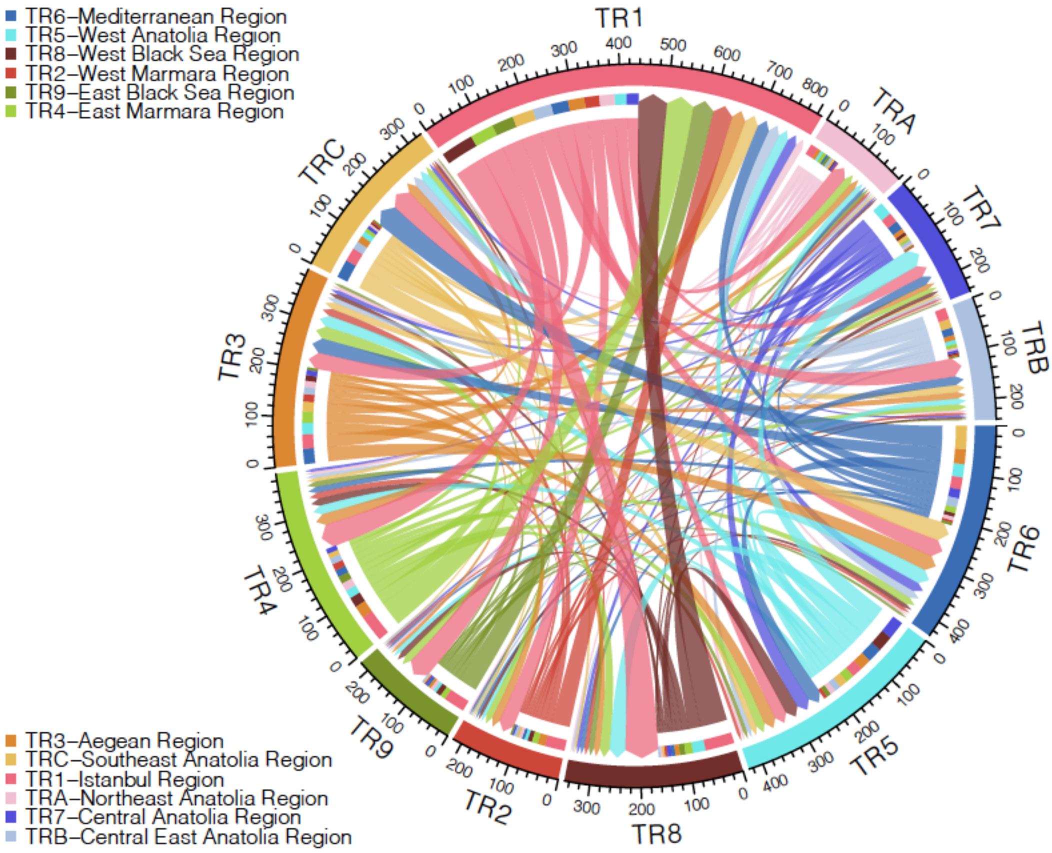

Figure 2.23: Bilateral migration flows at NUTS1 level in Turkey (thousands), 2013

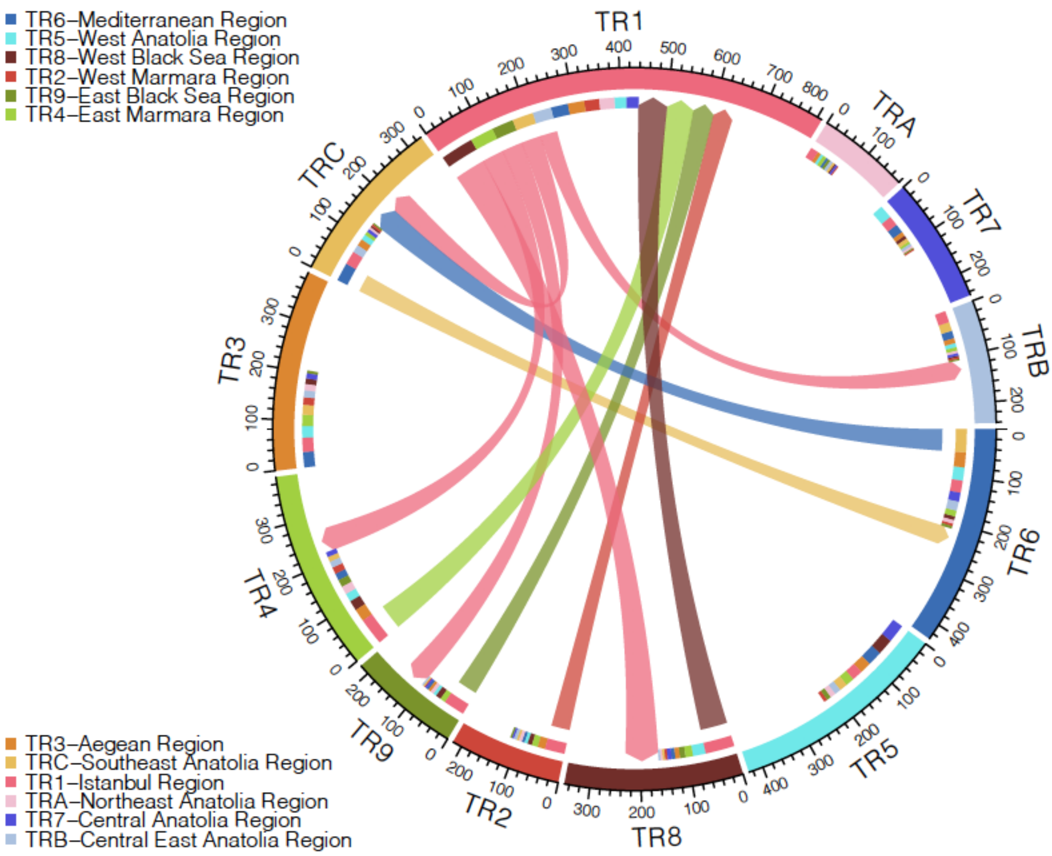

Figure 2.24: Top 10 largest bilateral migration flows at NUTS1 level in Turkey (thousands), 2013

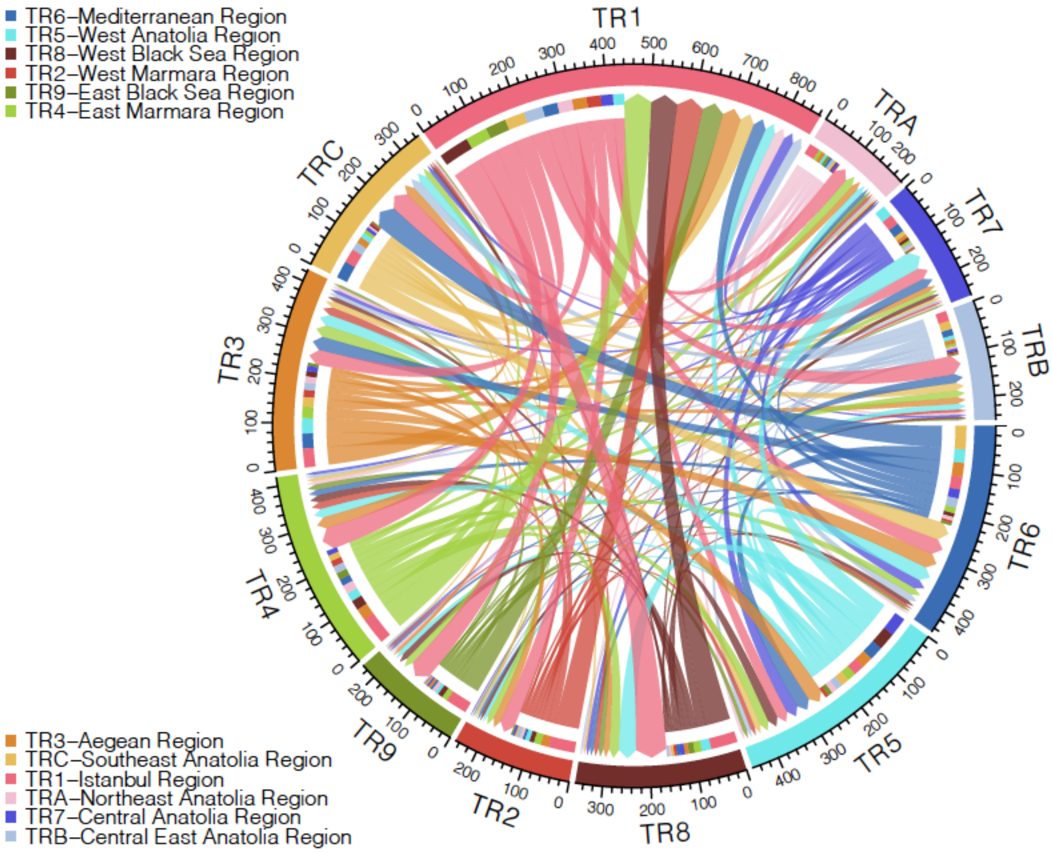

Figure 2.25: Bilateral migration flows at NUTS1 level in Turkey (thousands), 2014

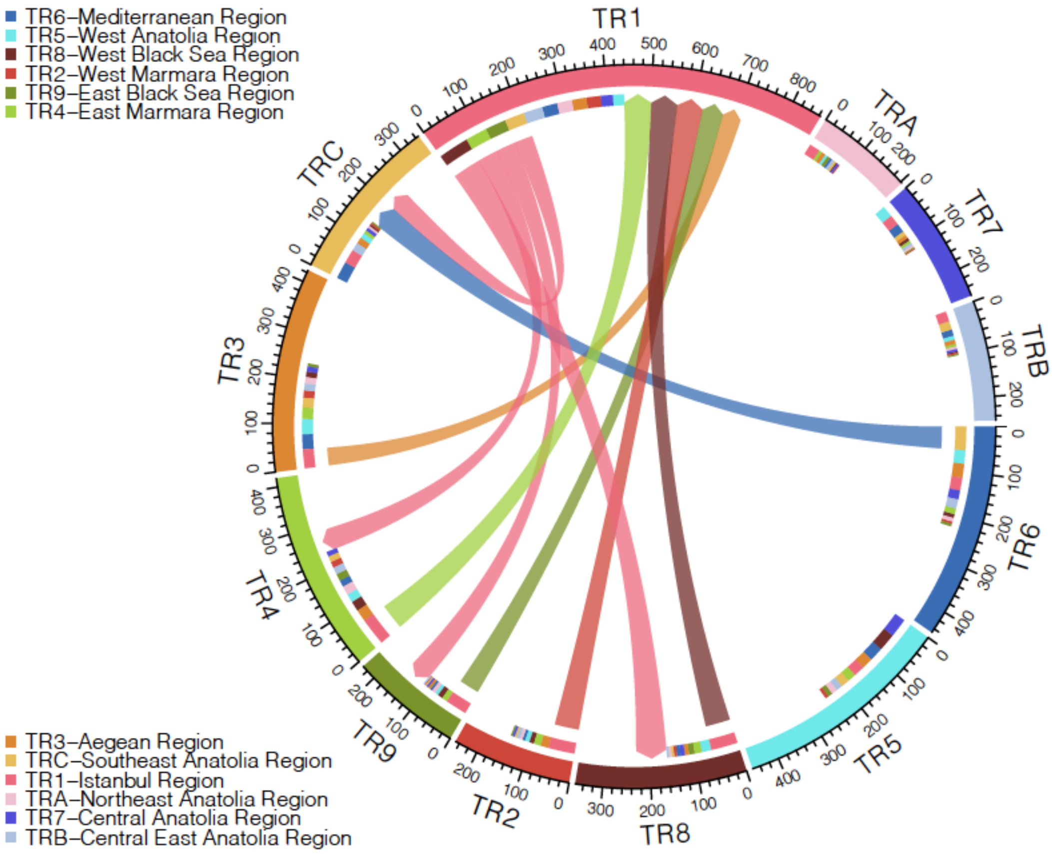

Figure 2.26: Top 10 largest bilateral migration flows at NUTS1 level in Turkey (thousands), 2014

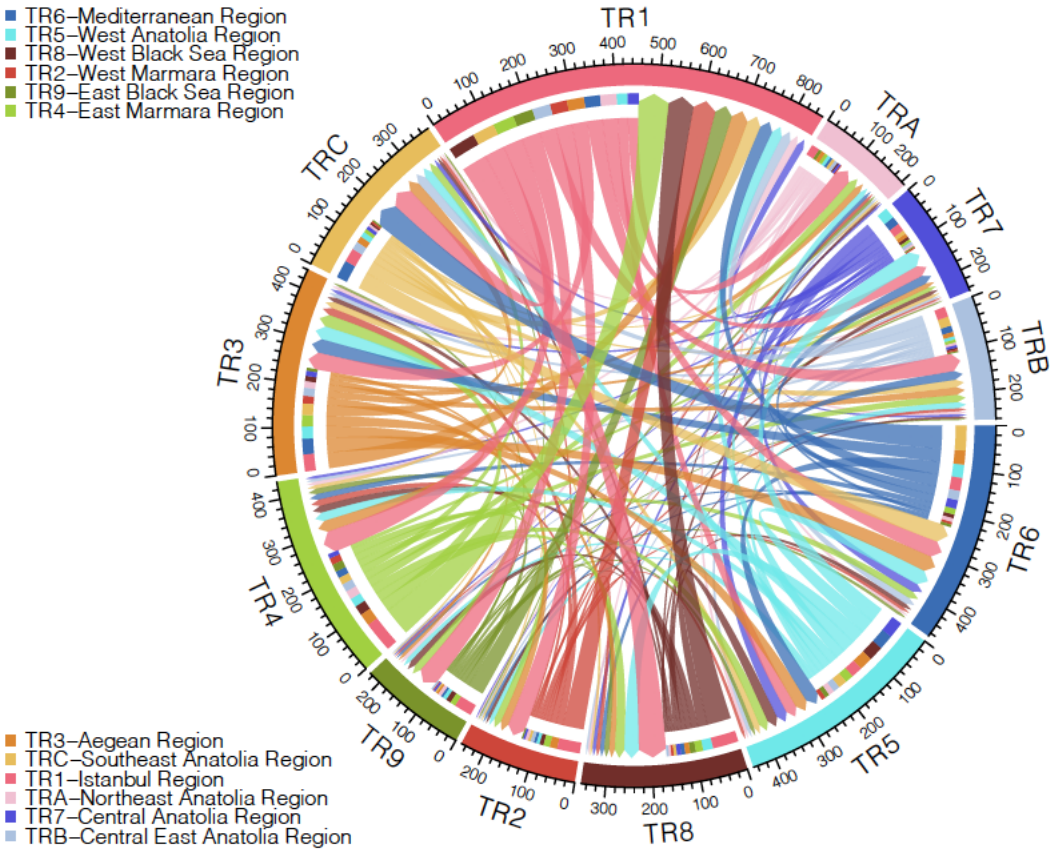

Figure 2.27: Bilateral migration flows at NUTS1 level in Turkey (thousands), 2015

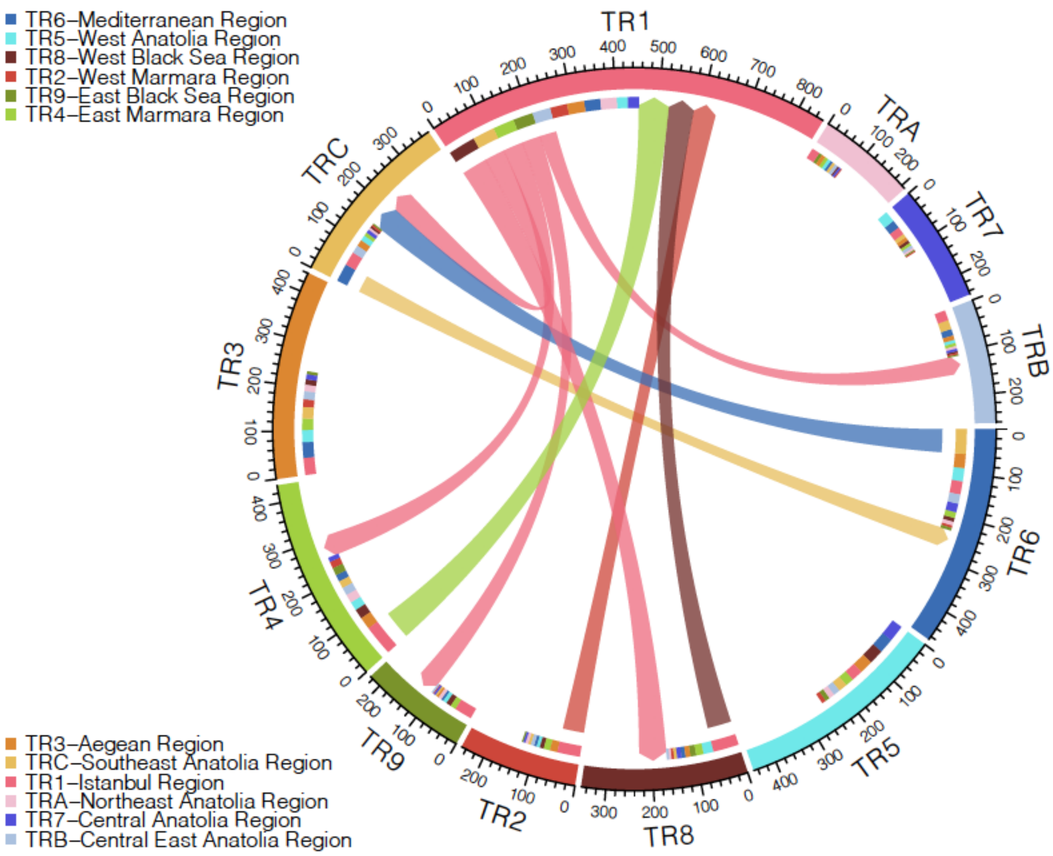

Figure 2.28: Top 10 largest bilateral migration flows at NUTS1 level in Turkey (thousands), 2015

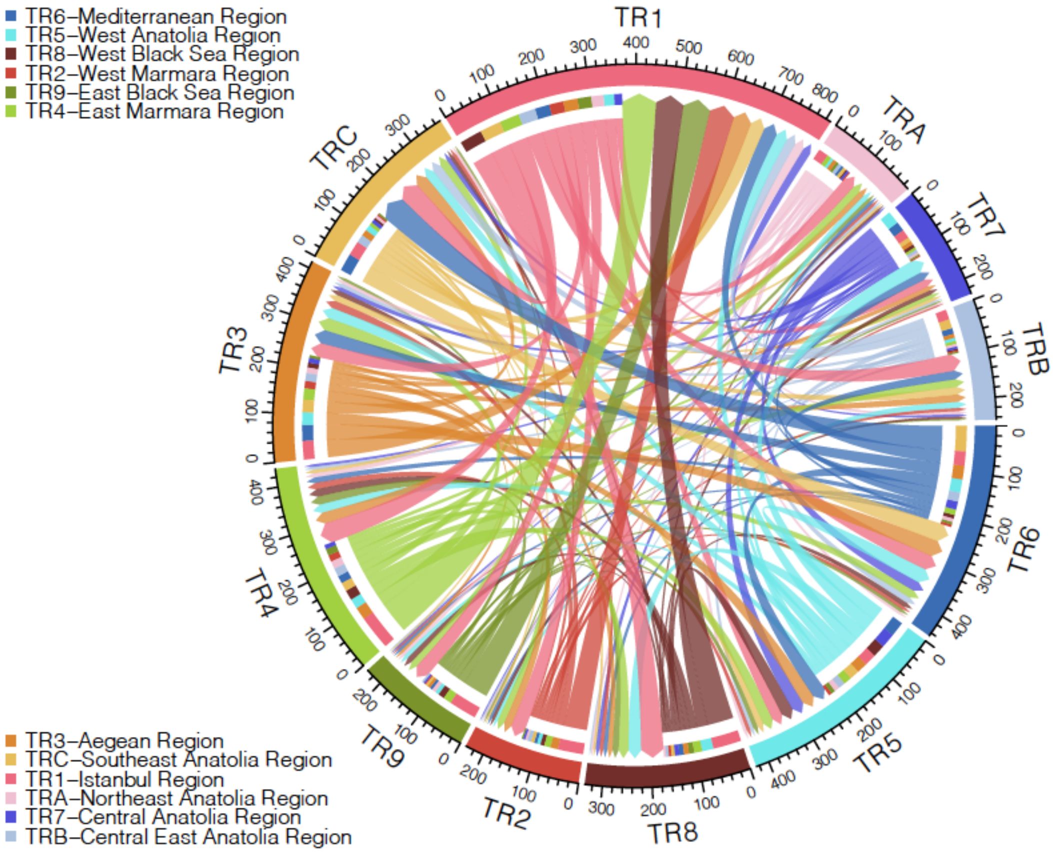

Figure 2.29: Bilateral migration flows at NUTS1 level in Turkey (thousands), 2016

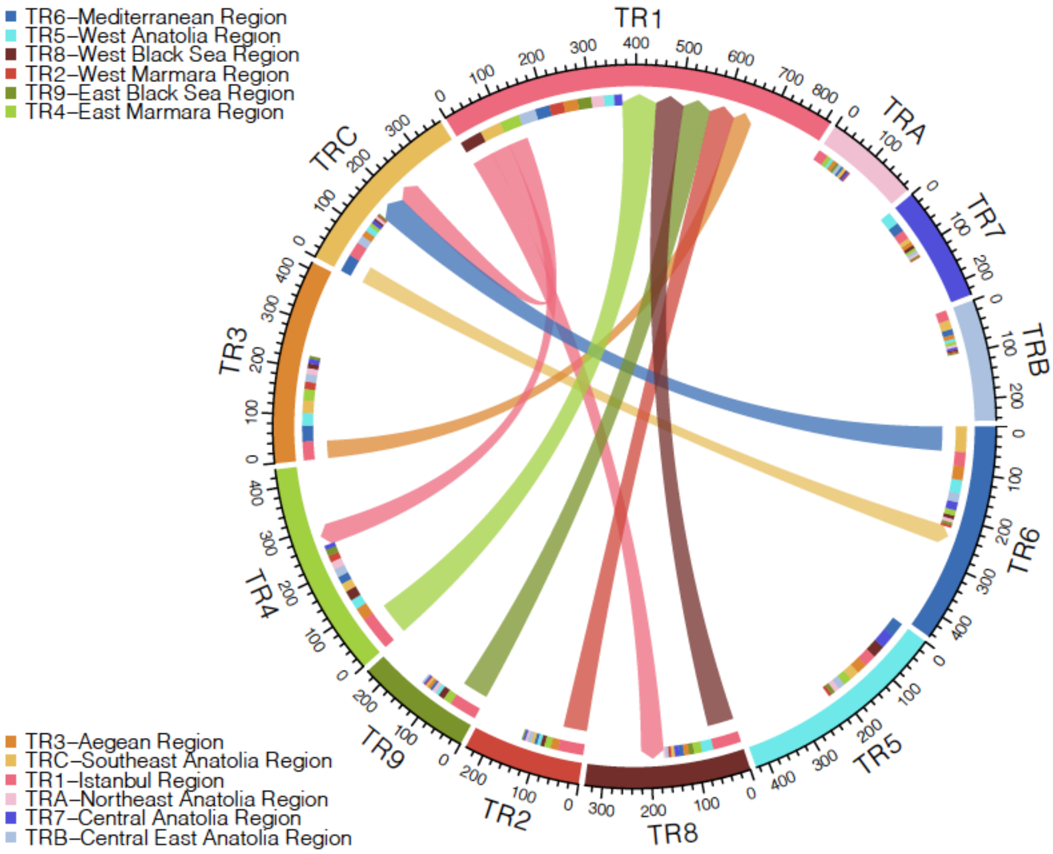

Figure 2.30: Top 10 largest bilateral migration flows at NUTS1 level in Turkey (thousands), 2016

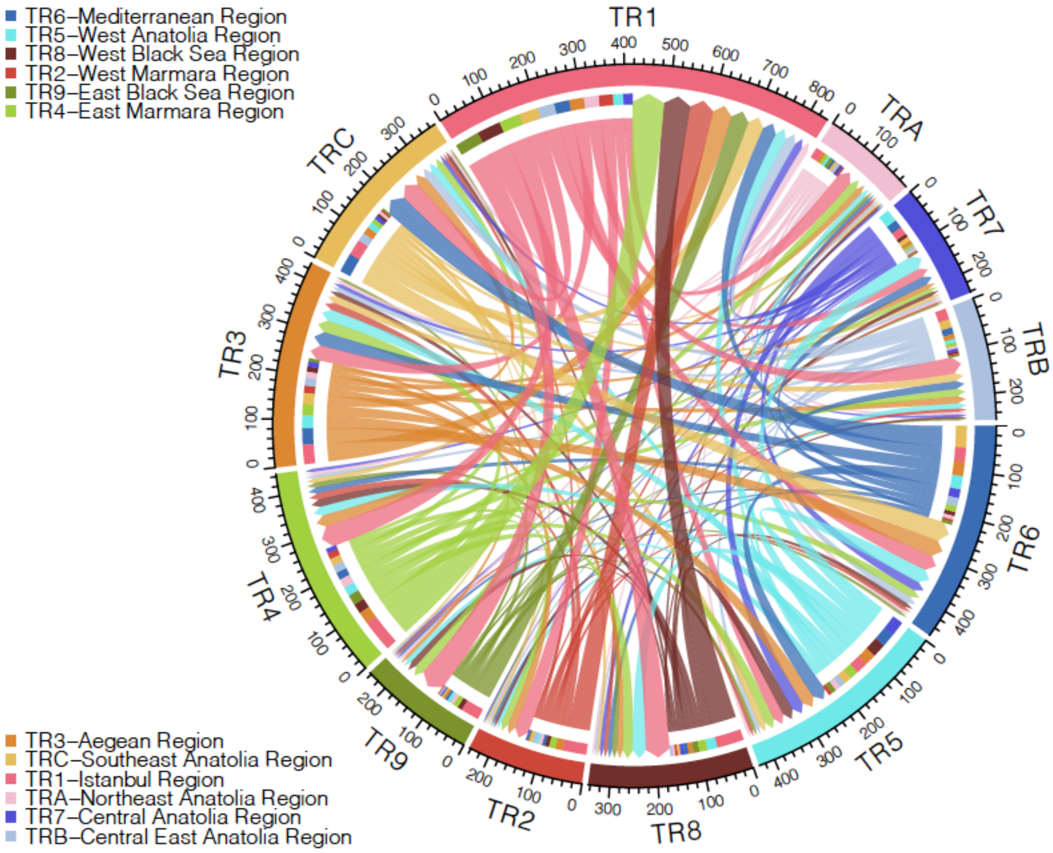

Figure 2.31: Bilateral migration flows at NUTS1 level in Turkey (thousands), 2017

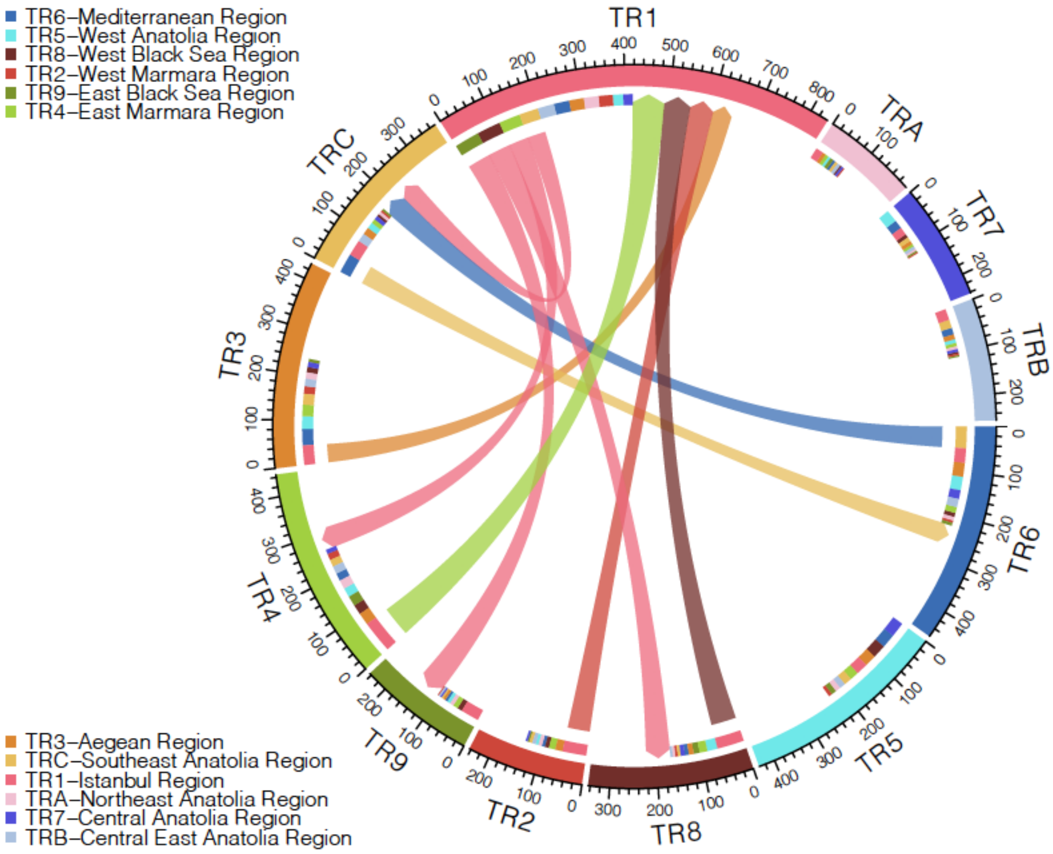

Figure 2.32: Top 10 largest bilateral migration flows at NUTS1 level in Turkey (thousands), 2017

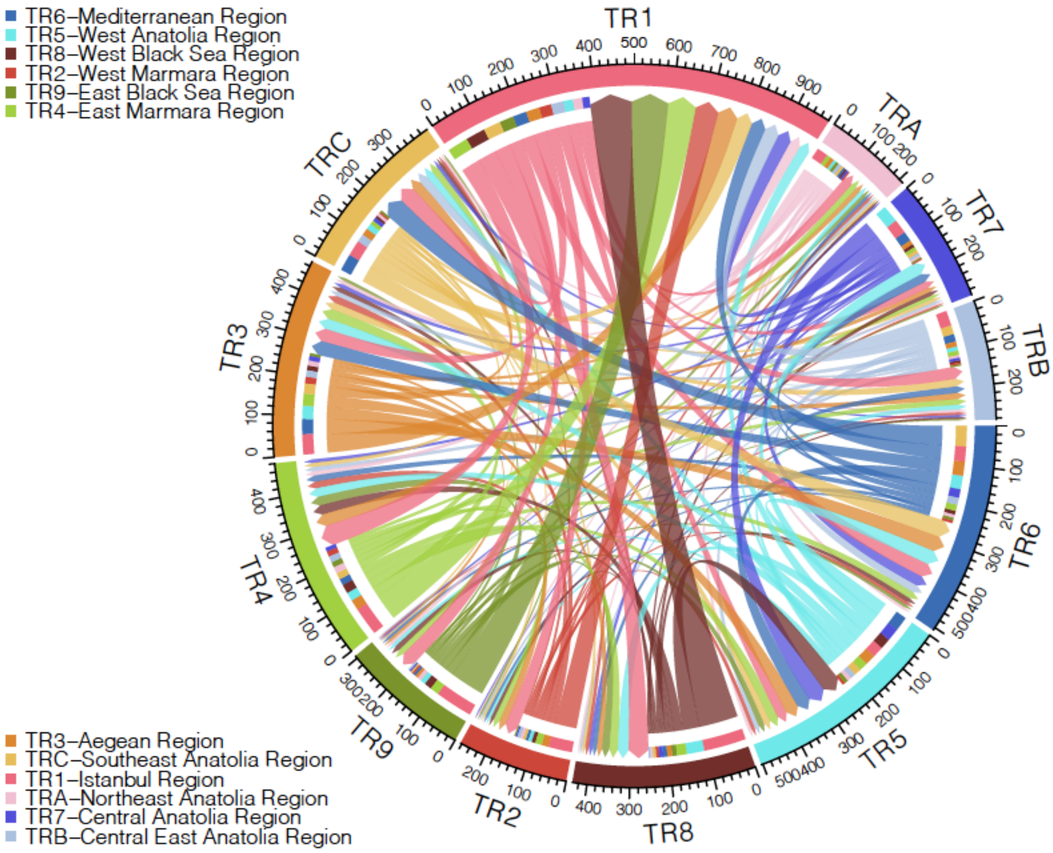

Figure 2.33: Bilateral migration flows at NUTS1 level in Turkey (thousands), 2018

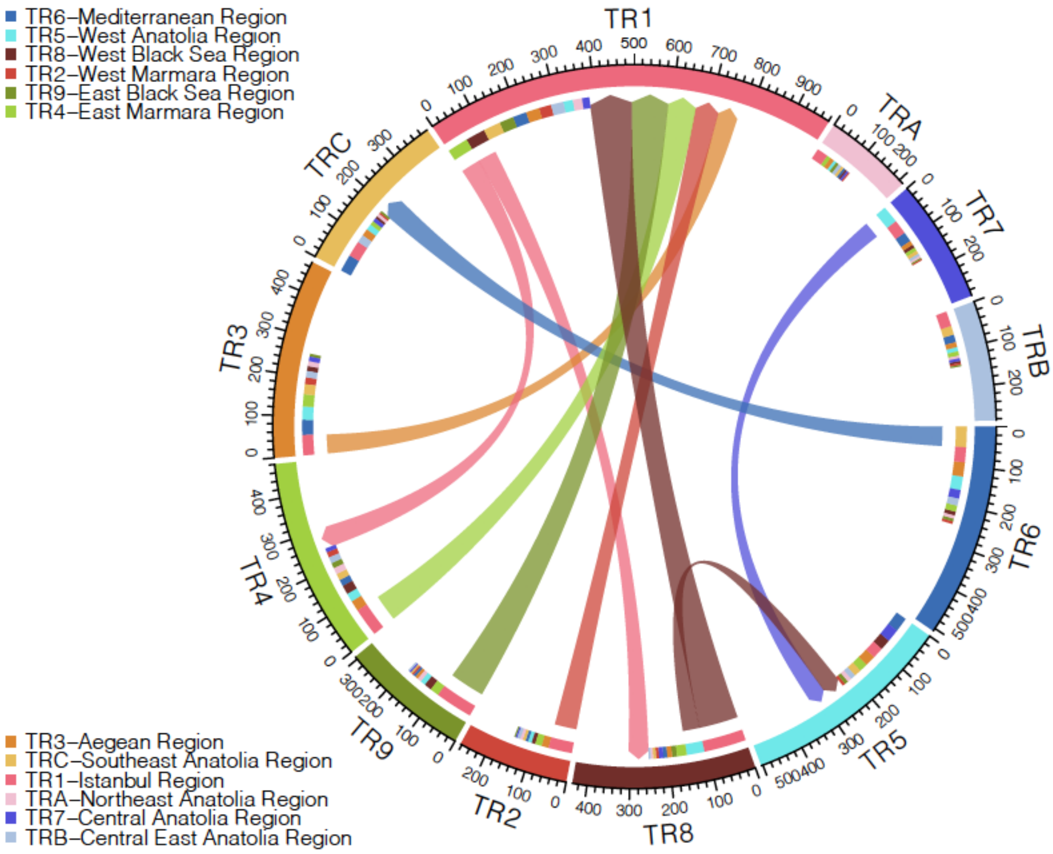

Figure 2.34: Top 10 largest bilateral migration flows at NUTS1 level in Turkey (thousands), 2018

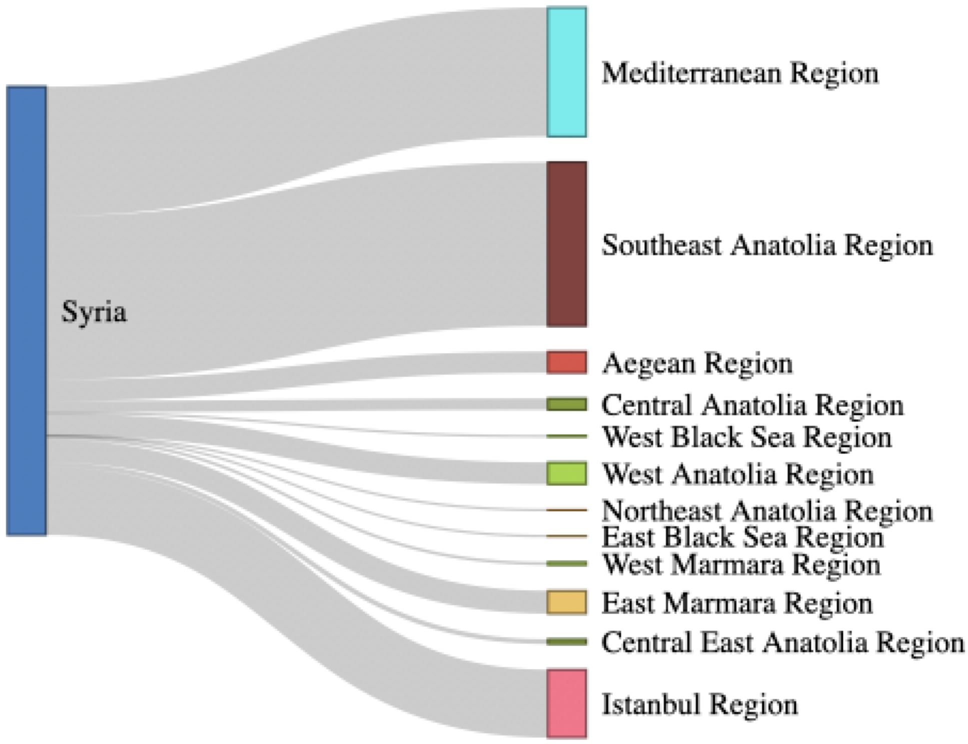

Figure 2.35: Distribution of SuTP at NUTS1 level (2016), Sankey diagram

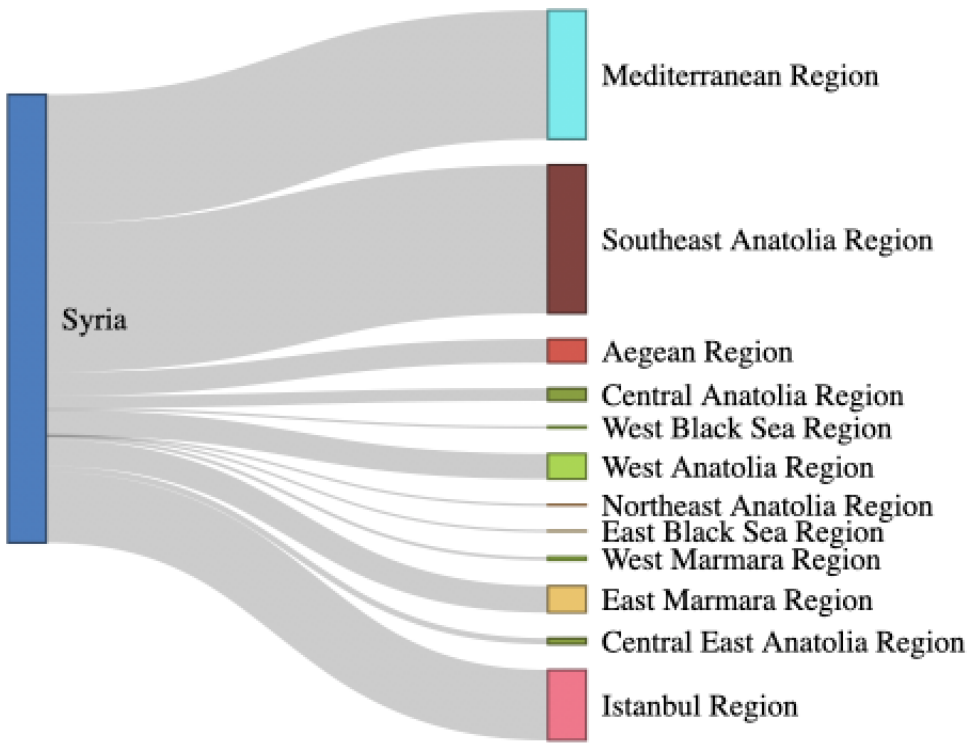

Figure 2.36: Distribution of SuTP at NUTS1 level (2017), Sankey diagram

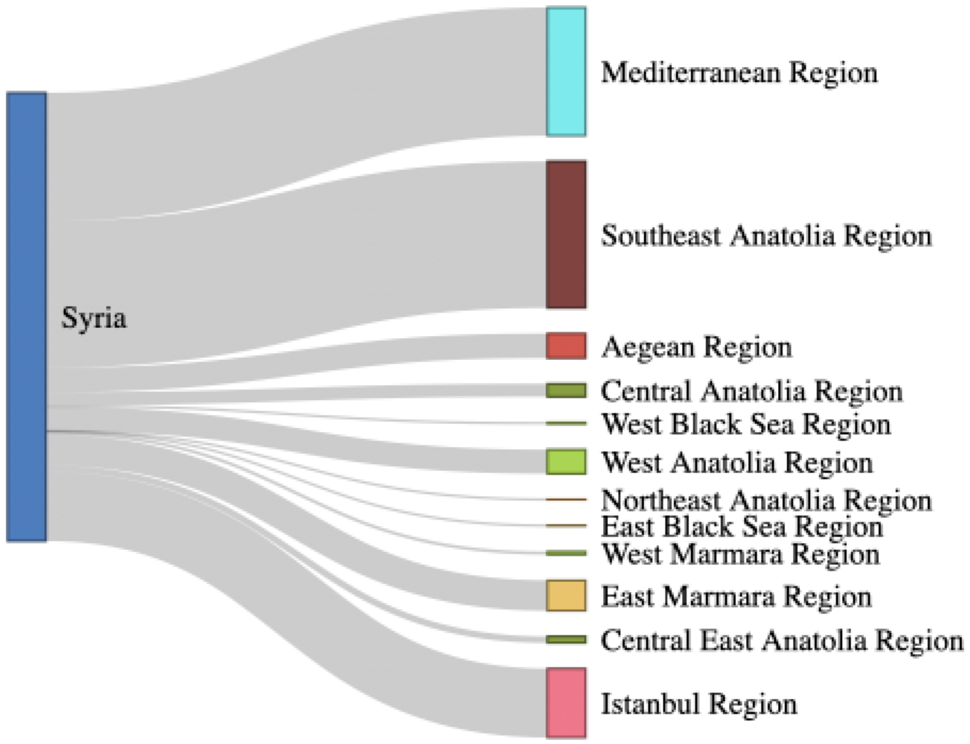

Figure 2.37: Distribution of SuTP at NUTS1 level (2018), Sankey diagram

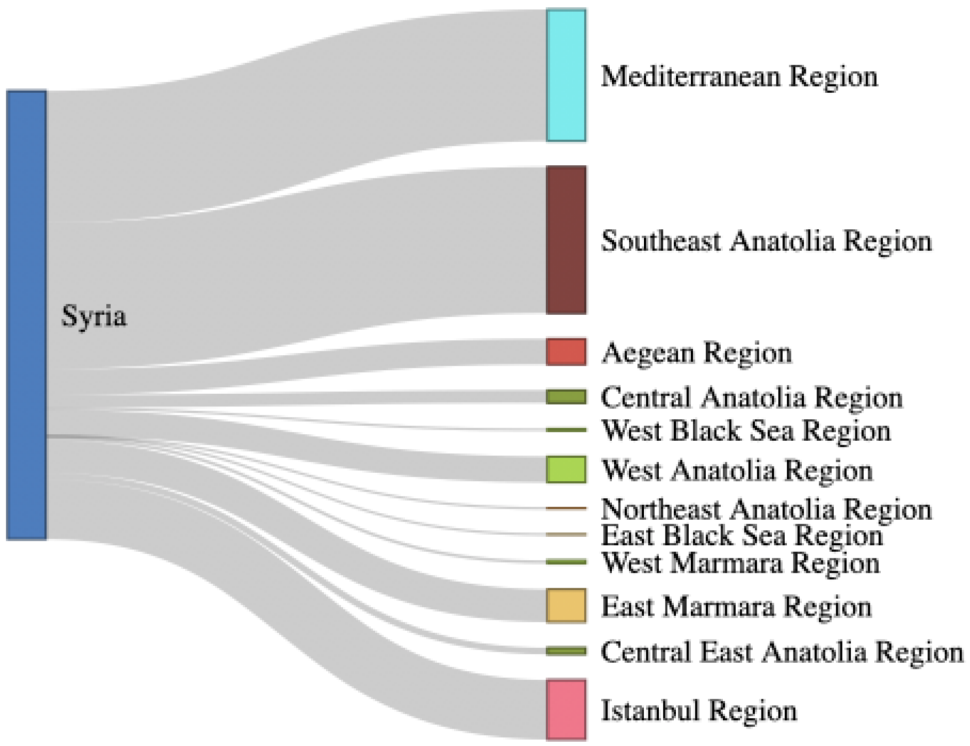

Figure 2.38: Distribution of SuTP at NUTS1 level (2019), Sankey diagram

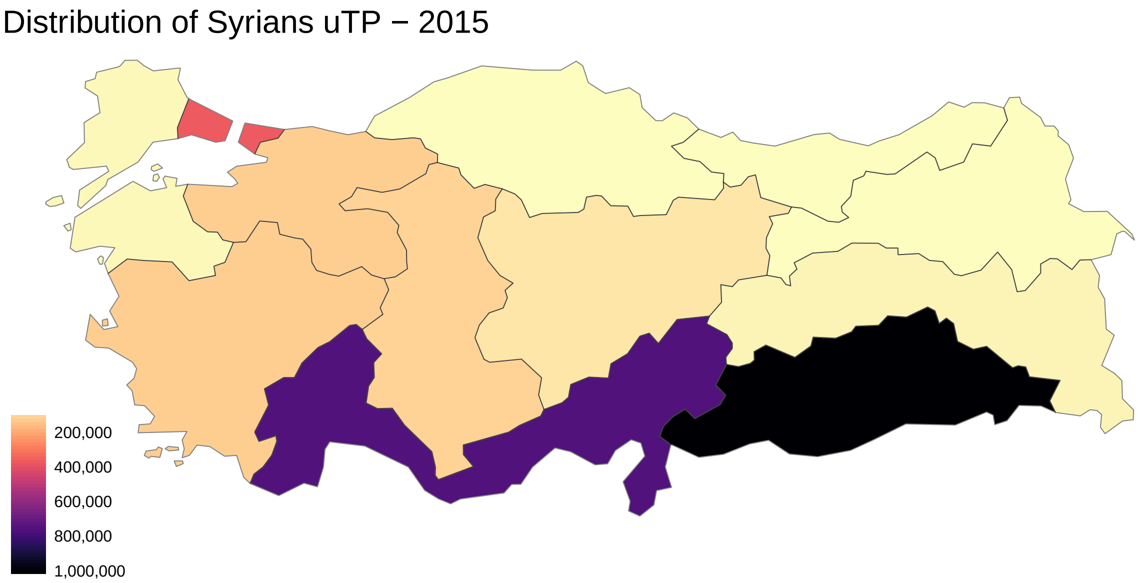

Figure 2.39: Distribution of SuTP at NUTS1 level (2015)

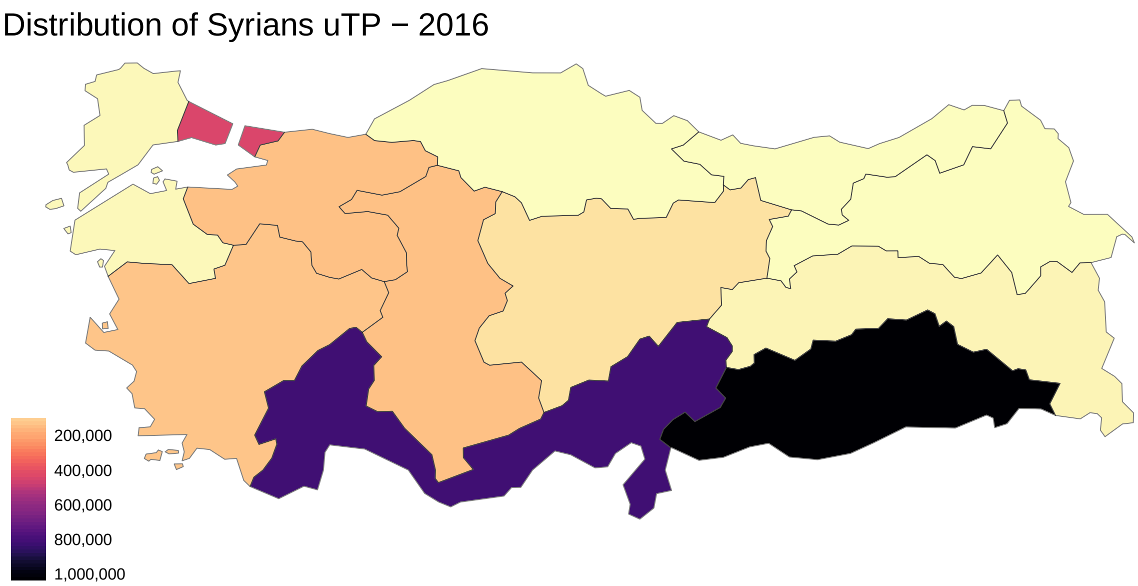

Figure 2.40: Distribution of SuTP at NUTS1 level (2016)

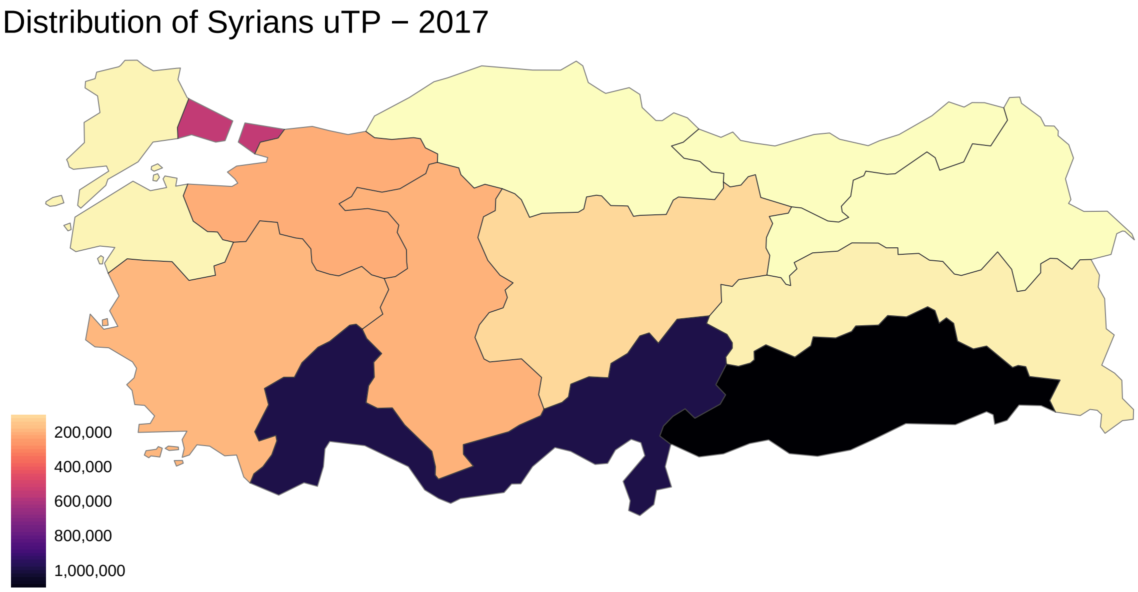

Figure 2.41: Distribution of SuTP at NUTS1 level (2017)

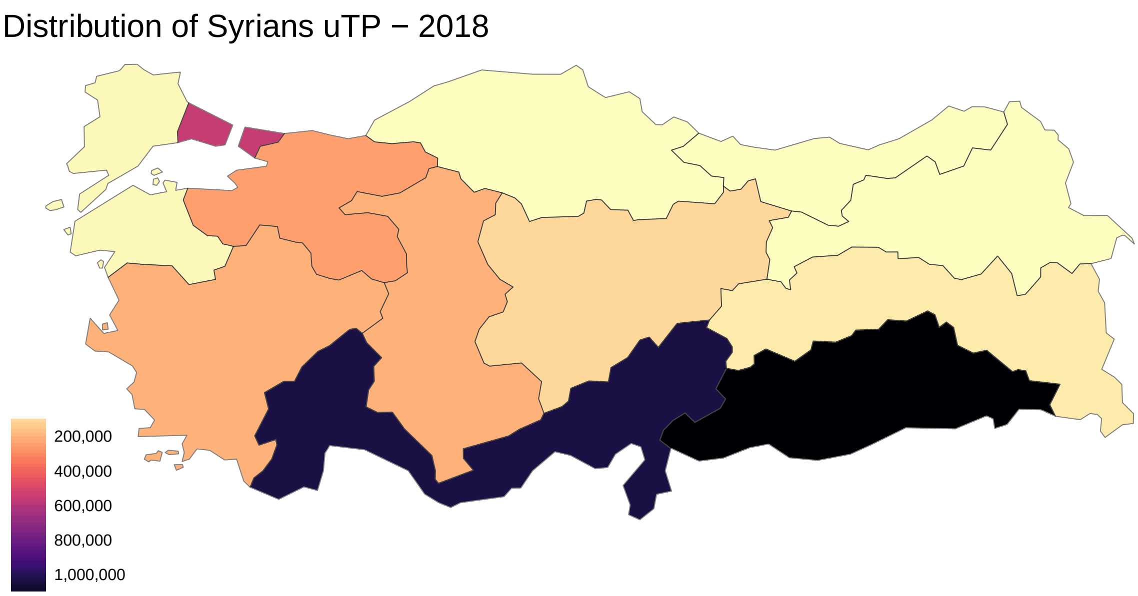

Figure 2.42: Distribution of SuTP at NUTS1 level (2018)

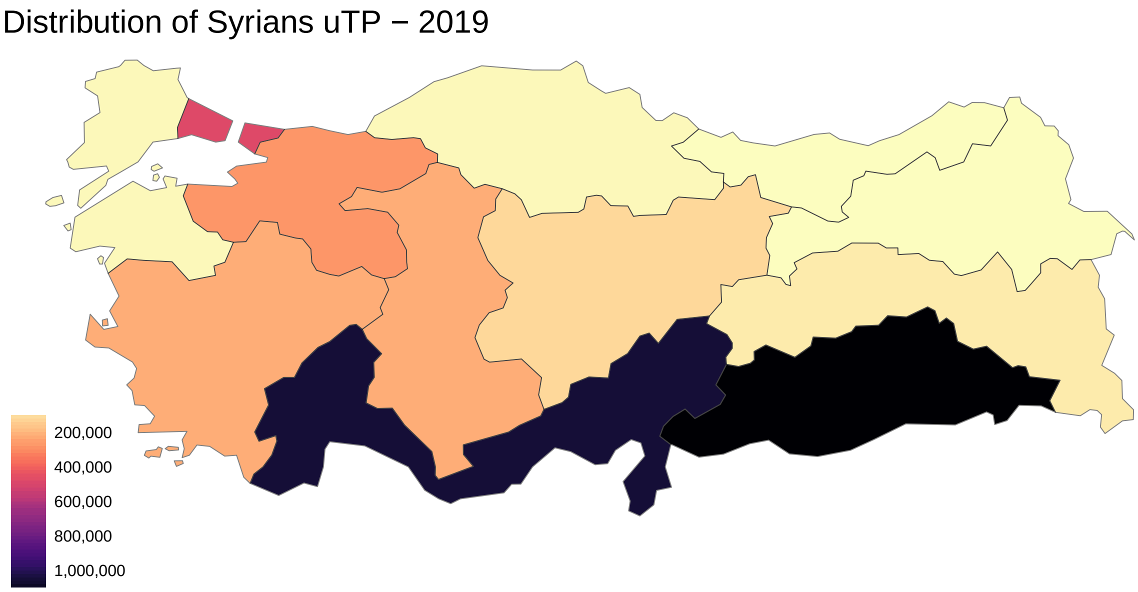

Figure 2.43: Distribution of SuTP at NUTS1 level (2019)

Figure 2.44: Distribution of native Arabic speakers in Turkey (1927 - # of people)

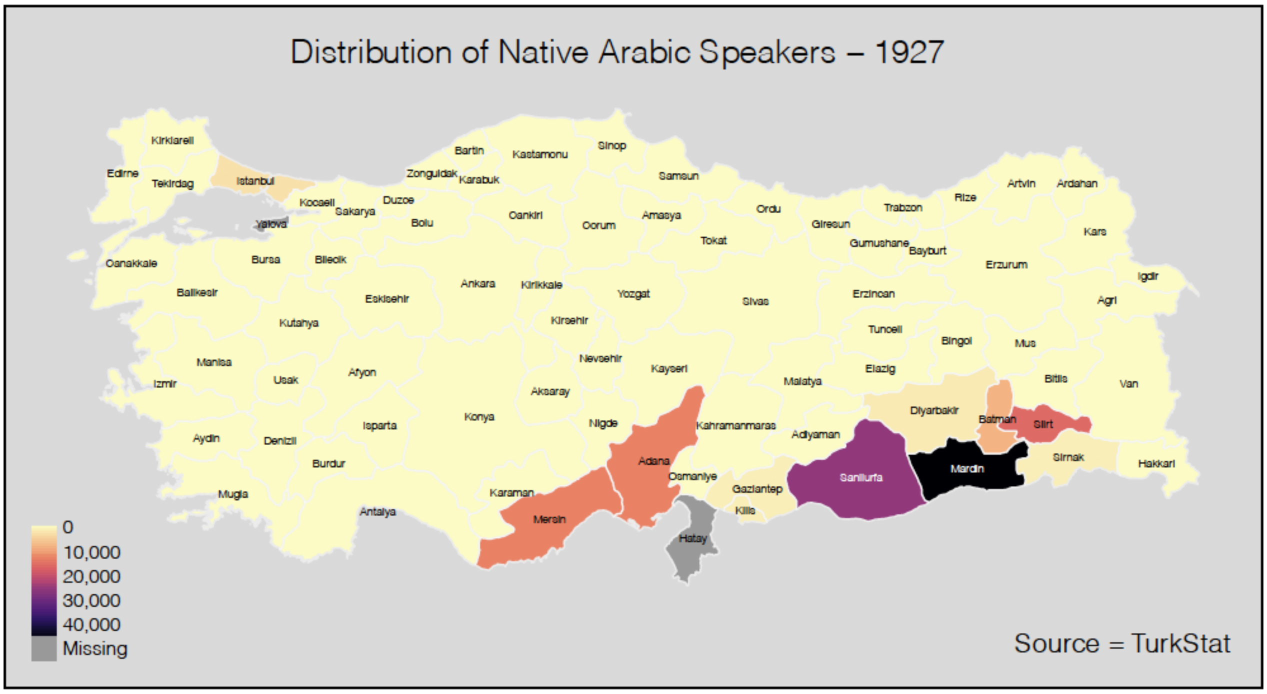

Figure 2.45: Share of native Arabic speakers in Turkey (1927 - complete data)

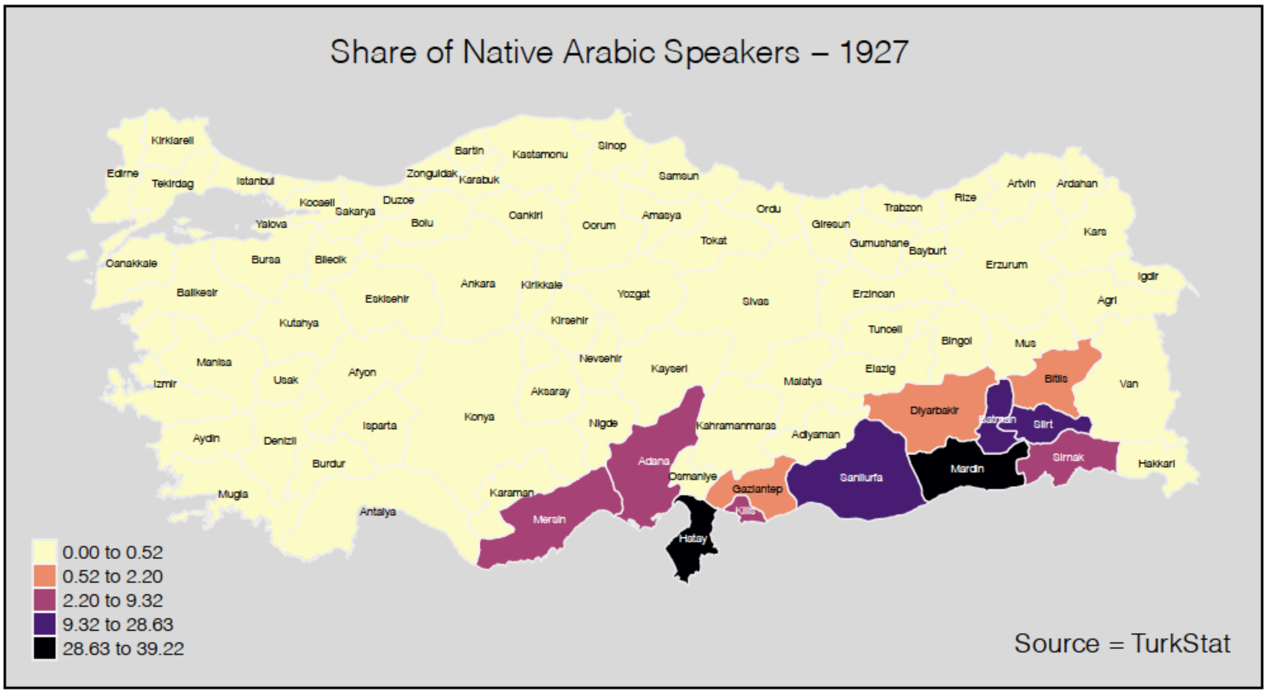

Table 2.4: Comparison of the results by random effects and fixed effects models

VARIABLES	(1) FE Model	(2) FE Model	(3) RE Model	(4) RE Model
Median Age	0.286*	-0.0517	0.426***	0.119
	(0.160)	(0.206)	(0.129)	(0.146)
Net Migration Rate	-0.00483***	-0.00174	-0.00197**	-0.00195**
	(0.00167)	(0.00145)	(0.000950)	(0.000814)
ln(Population)	7.424***	1.864	1.791***	1.800***
	(1.584)	(2.294)	(0.221)	(0.212)
Child Dependency Ratio	-0.0505	0.0251	0.0369	0.00904
	(0.0544)	(0.0461)	(0.0387)	(0.0371)
Elderly Dependency Ratio	-0.000834	-0.0465	-0.173*	-0.0667
	(0.140)	(0.166)	(0.0894)	(0.101)
Unemployment Rate	-0.0102	0.00461	0.0229**	0.00271
	(0.0121)	(0.0135)	(0.0109)	(0.0130)
Employment Rate	-0.0304***	-0.0170	-0.00572	-0.0179*
	(0.00877)	(0.0107)	(0.00669)	(0.0104)
Gini coefficient	-2.673	-1.994	-1.509	-1.595
	(1.719)	(1.882)	(1.766)	(1.748)
1927 Rate of Arabic Speakers	omitted	omitted	0.0205	0.0308**
			(0.0201)	(0.0148)
sqrt(Min. Distance)	omitted	omitted	-0.214***	-0.117***
			(0.0323)	(0.0383)
Constant	-94.99***	-14.10	-22.30***	-15.05***
	(20.51)	(33.47)	(3.515)	(3.689)
Observations	405	405	405	405
R-squared	0.425	0.517		
Number of provinces	81	81	81	81
Year FE	NO	YES	NO	YES
F-statistic	9.676	29.82		
Chi2			470.4	1134
Within R-squared	0.425	0.517	0.343	0.509
Between R-squared	0.437	0.623	0.727	0.737
Overall R-squared	0.430	0.620	0.716	0.730

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 2.5: Results of replications with alternative variables (First column reports the main model for comparison)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
VARIABLES	RE Model	RE Model	RE Model	RE Model	RE Model	RE Model	RE Model	RE Model	RE Model	RE Model
Net Migration	-0.00195**	-0.00195**	-0.00186**	-0.00194**	-0.00186**	-0.00185**	-0.00187**	-0.00185**	-0.00185**	-0.00175**
Rate	(0.000814)	(0.000817)	(0.000782)	(0.000818)	(0.000783)	(0.000786)	(0.000817)	(0.000813)	(0.000814)	(0.000759)
Median Age	0.119	0.119	0.0835	0.121	0.0772	0.0743	0.101	0.0904	0.0801	0.0442
	(0.146)	(0.145)	(0.151)	(0.148)	(0.151)	(0.152)	(0.147)	(0.149)	(0.152)	(0.0544)
ln(Population)	1.800***	1.814***	1.649***	1.832***	1.635***	1.646***	1.800***	1.748***	1.744***	1.911***
	(0.212)	(0.211)	(0.206)	(0.213)	(0.207)	(0.210)	(0.217)	(0.221)	(0.222)	(0.141)
Rate of Arabic	0.0308**				0.0261*	0.0266*	0.0412***	0.0531***	0.0571***	0.0323**
Speakers (1927)	(0.0148)				(0.0148)	(0.0148)	(0.0136)	(0.0130)	(0.0138)	(0.0158)
Unemployment	0.00271	0.00257	0.00826	0.00422	0.00662	0.00634	0.00299	0.00438	0.00480	0.00315
Rate	(0.0130)	(0.0130)	(0.0124)	(0.0128)	(0.0126)	(0.0126)	(0.0130)	(0.0132)	(0.0132)	(0.0129)
Employment	-0.0179*	-0.0179*	-0.0193*	-0.0183*	-0.0187*	-0.0186*	-0.0175*	-0.0182*	-0.0189*	-0.0175*
Rate	(0.0104)	(0.0104)	(0.0104)	(0.0104)	(0.0104)	(0.0105)	(0.0104)	(0.0103)	(0.0103)	(0.0105)
sqrt(Min. Dist.)	-0.117***	-0.121***	-0.127***	-0.133***	-0.117***	-0.119***				-0.116***
	(0.0383)	(0.0392)	(0.0273)	(0.0343)	(0.0308)	(0.0311)				(0.0386)
Child Dep.	0.00904	0.00772	0.0266	0.0108	0.0234	0.0221	0.0103	0.0170	0.0184	
Ratio	(0.0371)	(0.0367)	(0.0377)	(0.0370)	(0.0378)	(0.0378)	(0.0371)	(0.0373)	(0.0373)	
Elderly Dep.	-0.0667	-0.0648	-0.0903	-0.0633	-0.0879	-0.0839	-0.0535	-0.0658	-0.0695	
Ratio	(0.101)	(0.101)	(0.0999)	(0.104)	(0.0996)	(0.102)	(0.103)	(0.104)	(0.106)	
Gini coefficient	-1.595	-1.594	-1.538	-1.534	-1.597	-1.603	-1.650	-1.610	-1.621	-1.657
	(1.748)	(1.745)	(1.721)	(1.743)	(1.729)	(1.731)	(1.757)	(1.753)	(1.752)	(1.801)

Table 2.5 cont.

VARIABLES	(1) RE Model	(2) RE Model	(3) RE Model	(4) RE Model	(5) RE Model	(6) RE Model	(7) RE Model	(8) RE Model	(9) RE Model	(10) RE Model
Rate of Arabic		0.0337								
Speakers (1965)		(0.0252)								
Rate of Kurdish			-0.0242***		-0.0236***	-0.0230***				
Speakers (1927)			(0.00757)		(0.00792)	(0.00806)				
Rate of Circassian				0.0632		0.0359				
Speakers (1927)				(0.0617)		(0.0600)				
Minimum Duration							-0.221***			
(hrs)							(0.0688)			
Maximum Duration								-0.171***		
(hrs)								(0.0538)		
sqrt(Max. Distance)									-0.0798**	
									(0.0329)	
Dependency Ratio										-0.00683
										(0.0273)
Constant	-15.05***	-15.11***	-11.55**	-15.29***	-11.39**	-11.45**	-15.85***	-14.72***	-13.77***	-14.48***
	(3.689)	(3.670)	(4.632)	(3.630)	(4.675)	(4.672)	(3.873)	(4.029)	(4.080)	(3.491)
Observations	405	405	405	405	405	405	405	405	405	405
Number of provinces	81	81	81	81	81	81	81	81	81	81
Year FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Chi2	1134	1101	998.6	977.7	1079	1121	1016	939.5	819.5	1080
Within R-squared	0.509	0.509	0.513	0.509	0.513	0.513	0.511	0.512	0.513	0.512
Between R-squared	0.737	0.735	0.753	0.735	0.756	0.757	0.726	0.714	0.697	0.729
Overall R-squared	0.730	0.729	0.746	0.728	0.749	0.750	0.720	0.708	0.692	0.723

Robust standard errors in parentheses *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 2.6: Results of the replication excluding Istanbul

VARIABLES	(1) ln(Refugees)	(2) ln(Refugees)
Median Age	0.426***	0.119
	(0.132)	(0.148)
Net Migration Rate	-0.00192**	-0.00192**
	(0.000962)	(0.000824)
ln(Population)	1.777***	1.800***
	(0.250)	(0.237)
Child Dependency Ratio	0.0371	0.00929
	(0.0395)	(0.0376)
Elderly Dependency Ratio	-0.173*	-0.0672
	(0.0894)	(0.101)
Unemployment Rate	0.0227**	0.00254
	(0.0110)	(0.0130)
Employment Rate	-0.00533	-0.0175*
	(0.00681)	(0.0105)
Gini coefficient	-1.600	-1.490
	(1.812)	(1.806)
1927 Rate of Arabic Speakers	0.0207	0.0309**
	(0.0200)	(0.0148)
sqrt(Min. Distance)	-0.214***	-0.117***
	(0.0337)	(0.0400)
Constant	-22.08***	-15.11***
	(3.582)	(3.764)
Observations	400	400
Number of provinces	80	80
Year FE	NO	YES
Model	RE Model	RE Model
Chi2	434.6	1084
Within R-squared	0.342	0.509
Between R-squared	0.713	0.722
Overall R-squared	0.701	0.716

Robust standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 2.7: Replication of the model including GDP, (2019 removed)

VARIABLES	(1) ln(SuTP)	(2) ln(SuTP)
Median Age	0.474***	0.189
	(0.109)	(0.117)
Net Migration Rate	-0.00374***	-0.00177
	(0.00114)	(0.00110)
ln(Population)	1.647***	1.529***
	(0.210)	(0.200)
Child Dep. Ratio	0.0343	0.0361
	(0.0338)	(0.0312)
Elderly Dep. Ratio	-0.270***	-0.123
	(0.0795)	(0.0831)
Unemployment Rate	0.0408**	0.0172
	(0.0199)	(0.0179)
Employment Rate	0.0149	-0.00981
	(0.0185)	(0.0198)
ln(GDP)	-1.043**	1.105**
	(0.416)	(0.537)
1927 Rate of Arabic Speakers	0.0248	0.0248*
	(0.0187)	(0.0139)
sqrt(Min. Distance)	-0.188***	-0.138***
	(0.0372)	(0.0356)
Gini coefficient	-0.896	-0.769
	(1.827)	(1.820)
Constant	-13.16***	-24.06***
	(4.425)	(4.303)
Observations	324	324
Number of provinces	81	81
Year FE	NO	YES
Model	RE Model	RE Model
F-statistic	563.8	1151
Between R-squared	0.713	0.765
Overall R-squared	0.706	0.760

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Chapter 3

Search for a New Home: Refugee Stock and Google Search

Ebru Sanliturk [1], Francesco C. Billari [1,2]

1. Bocconi University
2. Carlo F. Dondena Centre for Research on Social Dynamics and Public Policy

Abstract

The trends of queries made on online search platforms are increasingly being used as an estimator by researchers in economics and social sciences. Following the assumption that trends of online queries may indicate intentions and help to predict human behaviour, this study addresses the general issue of analysing, nowcasting and predicting migrant decisions and aims to shed light to this issue through an analysis of Google searches in the case of Syrian refugees' mobility in Turkey. In that respect, the paper exploits the difference in the alphabet used by Turkish and Syrian citizens as the method of differentiation between locals and Syrian refugees. The paper then examines the relationship between Google search queries for province names in Turkey and the number of Syrians under the temporary protection (TP) status across provinces. Aiming to contribute to the literature forecasting and nowcasting using digital data, we conduct an empirical analysis for the relationship between Google search frequency and Syrian citizens' settlement and re-settlement behaviour across provinces in Turkey for the period January 2016 – December 2019. Our analysis suggests that there is a positive and significant association between the Google search frequency index for province names and the number of Syrians under temporary protection in the relevant province.

About a decade ago, social scientists began to consider online generated data as a new source for data and the analyses and interpretation of the digital footprints of individuals acquired growing interest. Online generated data came to be considered as helpful especially where formal data, such as official registries, census data or national surveys, is lacking and data collection is burdensome and limited due to sensitivity and confidentiality issues. Data on refugees often suffer for both the lack of official registries and data protection. Digital or online generated data analysis thus has a potential to contribute to refugee & migration studies. Assuming that online search queries may indicate intentions and help to predict human behaviour, we address the issue of nowcasting and forecasting migration decisions and aim to shed light to this issue through an analysis of online searches in the case of Syrian refugees' mobility in Turkey. In our empirical analysis, we exploit the alphabetical difference between the local and refugee population as a method of differentiation and observe the association between Google search data and number of Syrians under temporary protection (hereafter SuTP) across provinces.

This paper consists of the following four sections. The first section "Background" provides both a brief literature review to demonstrate the conceptual framework of using online generated data in social sciences and introduces the case of Turkey and SuTP. The second section introduces the data and methodology used in the analysis of this paper. The third section presents and discusses the results of the empirical analysis. Finally the conclusion part summarizes the results of the analysis of this paper, together with potential policy implications and possible venues of future research.

3.1 Background

3.1.1 The Digital Footprints of Population Processes

The concepts of digital data and big data, also referred as online generated data, internet data or web data, came to the attention of researchers in social sciences in the last decade and paved the way for a growing literature in various fields of social sciences. Being publicly available and easy to access, real-time data collection and opportunities to gain insight on hard-to-reach populations are the main advantages digital data offer. At the same time, digital data also pose challenges, such as the problem of representativeness due to lack of universal access to internet and smartphones (Cesare et al., 2018). In response to the challenge of representativeness, the concepts of online populations and digital divide emerged, addressed by different methodologies using online data sources such as Twitter (Yildiz et al., 2017) and Facebook (Gil-Clavel and Zagheni, 2019). Demography literature witnessed an increase in the use of digital data to analyse various types of demographic behaviour. Digital data contributed to the study of; human mobility via mobile phone data (Palmer et al., 2013), online dating markets (Bruch and Newman, 2018), assortative mating based on online dating behaviour (Thomas, 2020), and parenthood via Twitter data (Mencarini et al., 2019). Building on the new opportunities of access to hard-to-reach populations, mobile phone ownership and usage was found to be associated with the fertility transition in sub-Saharan Africa (Billari et al., 2020).

Migration also benefited from the increased use of online data in research, as this new source of data offered new insights where traditional data and official records were not sufficient. The geo-location information provided by most of social media platforms is used as a proxy to examine international migration. Thus, social media platforms such as Facebook (Zagheni et al., 2017a), Twitter (Hawelka et al., 2014) and LinkedIn (State et al., 2014) became important data sources to monitor and interpret migration flows.

The main source of digital data in this study, online search trends and query intensity (popularity), attracted the attention of researchers as a possible estimator to predict future tendencies and events (*forecasting*) as well as to interpret the present, defined as *now-casting*. Examples of such research first emerged in economics and epidemiology. In the field of economics, online search data is first used to estimate certain macroeconomic indicators, such as estimating unemployment rate by job searches (Ettredge et al., 2005; Askitas and Zimmermann, 2009), predicting economic activity such as sales in automotive industry and real estate by online search data (Choi and Varian, 2009) or estimating inflation rate by expected inflation measured by searches on Google (Guzman, 2011).

In the studies mentioned above, online search data is mostly conceived as an estimator of future tendencies and statistics, in other words, as a means of *forecasting*. In contrast, in the field of epidemiology, online search data is suggested as a measure to predict the present and/or immediate future. Deriving from the assumption that increased online search frequency for the early symptoms of contagious diseases indicate a risk of an outbreak, epidemiological research used online search data to *now-cast* infectious disease outbreaks such as influenza (Ginsberg et al., 2009), chickenpox (Pelat et al., 2009) and salmonella (Brownstein et al., 2009).

Online search data also contributed to the research in demography. Demographic literature provided evidence for an association between internet search patterns and; abortions (Reis and Brownstein, 2010), fertility (Billari et al., 2016), top causes of mortality in the US (Ricketts and Silva, 2017), suicides & self-injuries in the US (McCarthy, 2010), in Italy (Solano et al., 2016), in Korea (Song et al., 2014) and in Taiwan (Chang et al., 2015).

Regarding the topic of this research, tracking migration through Google Trends data, three studies should be mentioned. In order of increasing relevance, the first is by Lin, Cranshaw and Counts (2019), examining the domestic migration flows through search queries made on Bing.com. The study demonstrates a consistent and high correlation between the domestic migration predictions obtained through the analysis of search queries with demographic controls and official domestic migration records. This study is important as it successfully shows the migration demand and potential changes in this demand by analysing internet search frequencies, which can provide a leverage for urban policy-makers for better planning in destination cities/states. The second study is by Wladyka (2013), in which he uses Google search data as a predictor of migration flows from Argentina, Colombia and Peru to Spain. Similar to the approach of this research, Wladyka (2013) considers queries related to migration to Spain as an indicator of intention for resettlement as well as a *forecast* measure for migration and compares the search frequency with official migration records in Spain. His assumptions rely on the Benson-Rea and Rawlinson framework for migration, consisting of five stages; 1) premigration, 2) information-search and migration decision, 3) migration and arrival, 4) post-arrival and early settlement and 5) settlement outcome (migrate elsewhere, successful settlement and return home) (Benson-Rea and Rawlinson, 2003, p. 66). The third research, that in fact inspired this study, is published by Pew Research Center and examines the relationship between queries made on Google by Syrian refugees and their movements from Syria to their destination country. In this study, Connor (2017) exploits the language differences between Syrians (speaking Arabic) and the countries on their route to the destination country in 2015 up until the EU-Turkey Deal comes into effect in 2016. The analysis shows that there is a significant association between the increase in the Googles search intensity for the names of the countries en route to the destination country (Germany) in Arabic and the increase in the number of Syrian asylum seekers recorded in these countries.

This study is inspired by Connor's (2017) analysis on the case of Syrian refugees, which shows a sharp decrease in queries made with the intention to migrate to the EU following the introduction of EU-Turkey Deal in 2016. Moving from the end point of Connor's analysis regarding Turkey and changing his strategy to use differences in language to differences in alphabet, we aim to analyse whether a similar pattern exists for the settlement decisions of SuTP inside Turkey.

3.1.2 Syrians under Temporary Protection in Turkey

The breakout of civil war in Syria occurred in 2011 and the situation escalated, with the consequent refugee influx into Mediterranean countries in 2014-2015. Since 2011, Turkey applied an open-door policy to Syrian refugees. As the urgency of the crisis became evident, the Turkish government accepted to become a part of Syrian Regional Response Plan (SRRP) in cooperation with the UNCHR (Kirişci and Ferris, 2015) and signed the controversial readmission agreement with the EU (known as the EU-Turkey Deal) in 2015. The EU-Turkey Deal imposed Turkey to intercept refugees who wish to enter the EU borders through Turkey as well as strengthen its capabilities to register, accommodate and facilitate Syrian refugees, while it conditioned the EU to accept one Syrian refugee to be resettled in the EU for each Syrian refugee being returned to Turkey by the EU (known as the one-to-one rule). The EU-Turkey Deal is important for this research as it marks the starting point of the period analysed in this research, as the provincial-level data became publicly available after this date.

Understanding the settlement patterns of SuTP is important in the case of Turkey. Once the biometric registration of Syrian refugees after their arrival to the country is complete, they hold the status of temporary protection (hereafter TP), a status created to address the Syrian refugee crisis exclusively (Regulation, 2014). Less than 5% of the SuTP in Turkey live in camp environments, while the remaining majority lives in urban areas, seeking jobs. Although unregistered Syrian refugees certainly exist, the TP status enables access to free healthcare and education as well as other forms assistance by both national and international initiatives. The access to public services and assistance is only valid in the province of registration, therefore obtaining and maintaining the TP status is considerably beneficial. SuTP outside the camps are allowed move to different provinces, provided that they obtain a travel permit that can be issued for various reasons such as work, family visit, healthcare and education by the Provincial Directorate of Migration Management (PDMM). After they move to the new province they need to register at that province in order to continue to benefit from the rights and privileges of TP status. By the end of December 2019, the total number of Syrian refugees under the status of temporary protection is 3,576,370. The figures 3.1 and 3.2 below show the change in the distribution of Syrian refugees across provinces by depicting the situation at the beginning of 2016 and at the end of 2019 respectively, using (Ministry of Interior) Directorate General of Migration Management data.

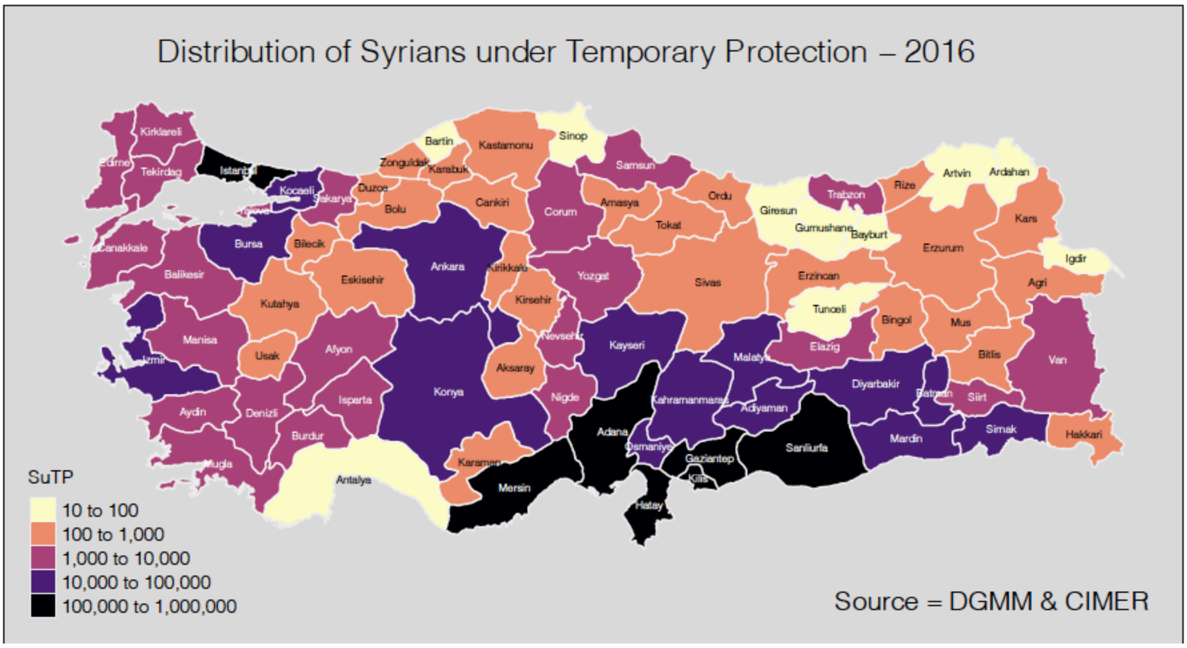

Figure 3.1: Distribution of Syrians under Temporary Protection across Provinces in Turkey, 2016

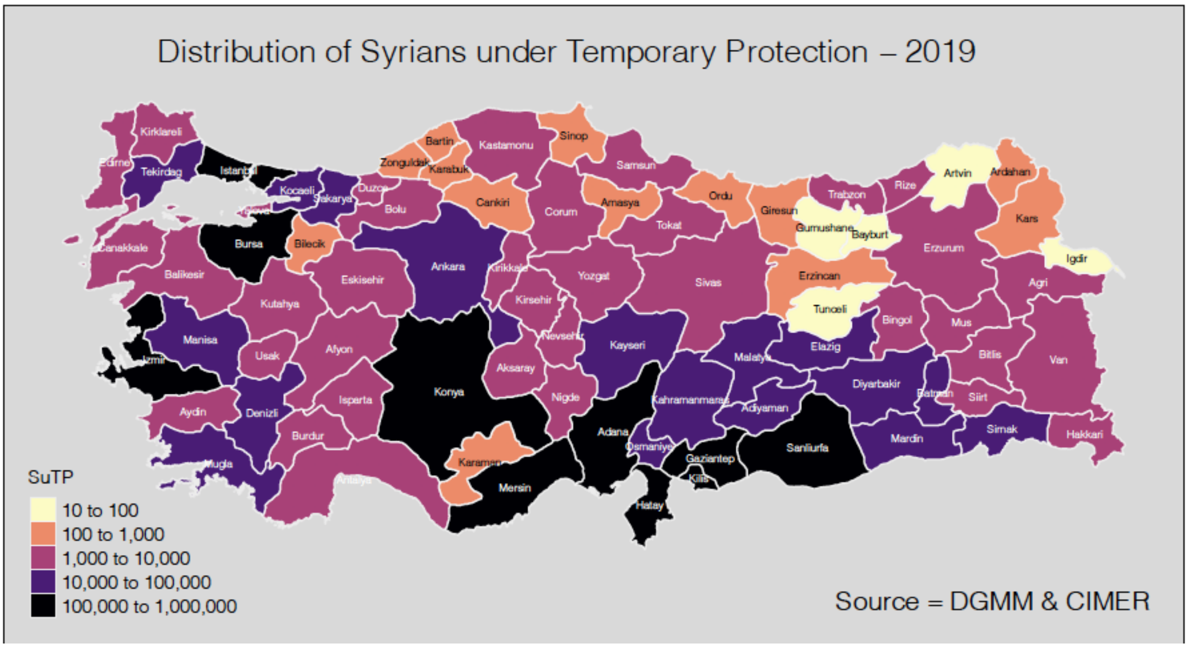

Figure 3.2: Distribution of Syrians under Temporary Protection across Provinces in Turkey, 2019

3.2 Data and Methods

3.2.1 Data

The period analysed in this study begins with the enactment of the EU-Turkey Deal in January 2016, when detailed statistics on Syrian refugees under temporary protection started to be published and regularly updated, and ends on December 31, 2019.

The empirical analysis in this study seeks to shed light to the any association between the official number of registered SuTP across provinces in Turkey (81 provinces in total) and internet search frequency for the names of in Turkey. The source of the data on the number of SuTP is the (Ministry of Interior) Directorate General of Migration Management[1] (hereafter DGMM), the public body responsible for the collection and publication of information on international migration. In line with the requirements of the EU-Turkey Deal, SuTP are registered at provincial level and the data is publicly available since January 15, 2016[2]. While the DGMM publishes weekly data on the registered SuTP by province, with each weekly update the preceding data disappears on the webpage. In order to obtain the data that was lost for public access on DGMM website, we first used the government channels[3] and requested the data on all updates for the three-year period of this study via the Directorate of Communications[4]. However, the response

[1]Original name of the institution is *(İçişleri Bakanlığı) Göç İdaresi Genel Müdürlüğü*

[2]Data on Syrian refugees under temporary protection was available only at country level before that date.

[3]Based on the rights recognized by the Law on the Right to Information (n. 4982) regulating requests for data & information from public institutions.

[4]Original name of the institution is *Presidency of the Republic of Turkey, Directorate of Communications*

to this request only included annual data, the number of SuTP across provinces at the end of 2016, 2017 and 2018. Needing an alternative method to obtain official data from the DGMM website in order to piece it together with the weekly data we managed to save in time and build a data set, we used the Internet Archive and its tool Wayback Machine. The Wayback Machine captures and archives webpages through web crawls, to save and retrieve the data lost for public access (Arora et al., 2016). Furthermore, we scanned the online news outlets (national and local) as well as the database of the National Thesis Center[5] as the distribution of SuTP across provinces is occasionally provided as descriptive statistics in theses and dissertations[6] on SuTP. Eventually we were able to build a data set of 122 weeks spanning from January 2016 to December 2019. The majority of our data for the number of SuTP was obtained through our own weekly downloads on DGMM website and via Wayback Machine.

The source of the internet search frequency for the names of Turkish provinces is Google Trends. Google Trends provides normalized data for the online search frequency of selected queries, rather than the sheer online search volume of the said queries. According to the guidelines provided by Google, Google Trends data is computed by dividing each data point by the total searches in the specified location and time range to compare relative popularity and the resulting numbers are then scaled on a range of 0 to 100 (Google, 2020). The weekly Google Trends data corresponds to the average of daily search frequencies in a week and is reported for each Sunday. In our study, the specified query index is the names of provinces in Turkey, in Arabic and Turkish letters[7]. The selected time period begins with November 15, 2015 and ends with December 31, 2019. The data for 2015 is used as the lagged search frequency and was not matched with the registry data of SuTP. Furthermore, for controls, we used Google Trends data for the same time period and query index in Syria, both in Arabic and Turkish letters as well as the international spelling of province names without special characters[8] in both Turkey and Syria.

Using the sources above, we created a unique panel data set covering four years, which includes 216 weeks' Google search data in six categories (Arabic letters in Turkey, Turkish letters in Turkey, English letters in Turkey, Arabic letters in Syria, Turkish letters in Syria and English letters in Syria) and 122 updates for the official number of SuTP in 81 provinces. On this data set we first calculated the main explanatory variable as the ratio of search frequencies for province names in Arabic letters over Turkish letters in Turkey ($\frac{search\ frequency\ in\ Arabic\ letters_{it}}{search\ frequency\ in\ Turkish\ letters_{it}}$), variable referred as the search frequency ratio, hereafter *SFR*. The reason for the use of SFR instead of separate search frequency scores for Arabic and Turkish spelling, is to absorb the effects of

(Cumhurbaşkanlığı İletişim Merkezi, CİMER)

[5]Full name; Council of Higher Education National Thesis Center (Yüksek Öğretim Kurumu Ulusal Tez Merkezi) accessible online via `https://tez.yok.gov.tr/UlusalTezMerkezi/giris.jsp`

[6]We included these descriptive statistics charts in our data sample only if they properly cited the same DGMM webpage source and only if the week of the update is specified.

[7]Turkish letters refers to Latin alphabet with the special characters in Turkish language (ç, ğ, ı, ö, ş, ü).

[8]The international spelling refers to the use of English language characters instead of special characters (c, g, i, o, s, u respectively).

potential seasonal popularity of provinces. As the SFR variable refers to the searches made in the same week of the update in the official registries for SuTP[9], this variable is considered to account for nowcasting estimations in our model. To account for forecasting, we calculated the 6 lagged variables with one-week intervals and for both the SFR variable and the search frequencies for controls over the period of 216 weeks. Once all the search frequency variables were set, we matched the dates of 122 updates of SuTP registry data with the date of search frequency variables by week. We then dropped the weeks that we could not match with the data on SuTP and this way we ensured that time-lags are not lost or mismatched due to the shrinking sample size.

3.2.2 Methods

The main aim of this study is to assess the relationship between online search frequency and mobility of SuTP across provinces in Turkey, considering the internet search behaviour as a sign of interest and intention to move. We seek to understand how Google search data can be used to nowcast and forecast in this case study, thus to analyse the relationship between simultaneous and lagged online searches and the observed mobility of SuTP. Deriving from these aims and assumptions, we define our hypotheses as (*H1*) online search data frequency is associated with the change in the number of SuTP in a given province and (*H2*) timing of the online search frequency plays a role to observe the association between online search behaviour and observed mobility. To test these hypotheses, we use our unique panel data set and a fixed effects model, considering the advantages of fixed effects model to absorb the effect of features that vary at province-level and time-level and allow a better interpretation of the explanatory variables used in this analysis. Thus, the baseline model is as follows;

$$ln(NumberSuTP)_{i[tj]} = \beta_1 \frac{search\ frequency\ in\ Arabic\ letters_{i[tj]}}{search\ frequency\ in\ Turkish\ letters_{i[tj]}} + \alpha_i + e_{i[tj]} \qquad (3.1)$$

In the specification above, the dependent variable is the natural logarithm of the number of SuTP in province i, at time tj, and the explanatory variable is the search frequency ratio (SFR) at time tj. We denote the time of the variable as tj, where j indicates the week of Google Trends data and t is the week in the registry data of SuTP. Thus the time denotation of tj refers to any week, for which Google search frequency and registry data of SuTP were matched (122 in total). The province-level fixed effects are denoted by α_i. The search query ratio refers to Google search frequency in Arabic letters for province i, at time t over Google search frequency in Turkish letters for province i, at time t, in Turkey. Note that the aim of using SFR is to control for potential local and seasonal effects; i.e. exceptional events at a certain time that may increase the interest in one province for both Turkish and Arabic speakers. Using the ratio allows us to observe the effect when search queries in one language changes more than the other.

[9] The official statistics of SuTP are published on weekdays whereas Google Trends reports weekly data for Sundays. Therefore the two sources of data are matched by week and not by day.

Next, in order to observe the predictive power of time-lags, that is the difference between the time of the Google search and the time refugee-mobility appearing on official data, we expand our model and add the lagged variables. While *tj* refers to the week for which the number of SuTP is matched with the same week of Google search query data, we denote the lagged variables as *tj-1*, *tj-4* and *tj-6*. In our data, each lag refers to a one-week interval and we introduce three lagged variables to the model; one-week, one-month and six-week lags. Following the addition of lagged variables, our model becomes as follows;

$$\begin{aligned} ln(NumberSuTP)_{i[tj]} =& \beta_1 \frac{search\ frequency\ in\ Arabic\ letters_{i[tj]}}{search\ frequency\ in\ Turkish\ letters_{i[tj]}} + \\ & \beta_2 \frac{search\ frequency\ in\ Arabic\ letters_{i[tj\text{-}1]}}{search\ frequency\ in\ Turkish\ letters_{i[tj\text{-}1]}} + \\ & \beta_3 \frac{search\ frequency\ in\ Arabic\ letters_{i[tj\text{-}4]}}{search\ frequency\ in\ Turkish\ letters_{i[tj\text{-}4]}} + \\ & \beta_4 \frac{search\ frequency\ in\ Arabic\ letters_{i[tj\text{-}6]}}{search\ frequency\ in\ Turkish\ letters_{i[tj\text{-}6]}} + \alpha_i + e_{i[tj]} \end{aligned} \tag{3.2}$$

Using online generated data as an explanatory variable raises questions about selection bias as the sample population is restricted to the portion of the population with internet access and high internet usage. To address this issue of selection bias, Zagheni and Weber (2015) propose to correct the estimates by internet penetration rate. This solution is problematic in the specific case of this study as the internet penetration rate in Turkey, where overall 83.8% of households have internet access (Institute, 2018), may not be representative for the SuTP. Thus, in order to avoid introducing a further bias, rate of internet access is not added to the model. Regarding the internet access rates of Syrians in Turkey, we rely on the assumption of high smart phone and internet usage for being considered as a crucial need by refugees (The GSM Association, 2017; Ulutürk et al., 2019) and on the survey demonstrating that 78% of the Syrians in Turkey own a smart phone and 36% consider internet/social media as the primary source of information (Sunata, 2017), while acknowledging the possibility of selection bias.

3.3 Results

The association between the online search frequency for province names in Turkey and number of SuTP across provinces are shown in Table 3.1. In all cases, robust standard errors are clustered at province level.

Table 3.1: Summary of results

	(1)	(2)	(3)	(4)
	ln(#SuTP)	ln(#SuTP)	ln(#SuTP)	ln(#SuTP)
$Arabic/Turkish\ SFR_{\ tj}$	0.0928***	0.0673***	0.0245**	0.00241
	(0.0236)	(0.0151)	(0.0111)	(0.0200)
$Arabic/Turkish\ SFR_{\ tj-1}$		0.0607***	0.0206*	0.0194*
		(0.0164)	(0.0109)	(0.0110)
$Arabic/Turkish\ SFR_{\ tj-4}$		0.0658***	0.0231**	0.0231**
		(0.0137)	(0.0104)	(0.0103)
$Arabic/Turkish\ SFR_{\ tj-6}$		0.0749***	0.0261**	0.0256**
		(0.0166)	(0.0127)	(0.0126)
ln(Population)			6.406***	6.286***
			(1.207)	(1.194)
No SF in Turkish Letters (SY)				-0.0587***
				(0.0191)
No SF in Arabic Letters (SY)				-0.00495
				(0.0142)
No SF in Arabic Letters (TR)				-0.0363
				(0.0345)
Constant	8.168***	8.108***	-76.71***	-75.04***
	(0.00813)	(0.0196)	(16.00)	(15.81)
Observations	9,882	9,882	9,882	9,882
R-squared (within)	0.014	0.041	0.207	0.211
Overall R-squared	0.0881	0.182	0.563	0.564
Number of Provinces	81	81	81	81
FE	YES	YES	YES	YES

Robust standard errors in parentheses ***
p<0.01, ** p<0.05, * p<0.1

In the table 3.1, the dependent variable is the natural logarithm of the number of SuTP in each province. In the first column, the explanatory variable is the Arabic/Turkish SFR for searches made on the same week as the date of the SuTP registry (denoted as tj) along with province-level fixed effects. In the second column one-week, one-month and six-week lags for the Arabic/Turkish SFR in Turkey are added to the equation, as indicated in Model 2 above. In the third and fourth columns controls are included to the regression equation; province-level population and dummy variables for the overall absence of online search frequency for province i at time t. for as controls respectively. As observed in the first two columns, the simultaneous SFR variable becomes insignificant with the addition of control variables and the analysis shows no support for a correlation between simultaneous online SF and SuTP registry. However, the analysis supports our *H1* and *H2* and confirms that there is an association between the number of SuTP in a province and the online SF for the name of the same province as well as the timing of the online SF matters. All three lagged SFR variables are observed to be statistically significant and positively correlated with the number of SuTP in the relevant provinces. We also observe that the less recent the lagged SFR variable is, the stronger is the correlation. The results suggest that an increase by one in the one-week prior SFR is correlated with a 1.9% increase in the number of SuTP. The association becomes stronger when the time gap increases, as one-month and six-week time lags are associated with 2.6% and 2.9% increase in the number of SuTP respectively. An increase of 2-3% in the number of SuTP in a given province in correlation with an increase by one in SFR should, however be considered carefully. As can be seen in Appendix (3.5), Table 3.2, the overall range of the SFR variable is 0 to 7.08 while the overall mean is 0.35. Therefore, an increase by one is an unlikely event and it is more likely that the association between the SFR and number of SuTP in a province occurs in much smaller amounts. Controls for the absence of SF is only significant for the absence of SF in Turkish letters in Syria. The lack of SF in Turkish letters in Syria is correlated with a 5.9% decrease in the number of SuTP. The result is intuitive as online searches for the provinces of another country using a foreign alphabet and special characters would be the most clear sign of interest.

To control for the strength of our results, we implemented a series of robustness checks. First we replicated the model using a random-effects approach, results of which are presented in comparison with the fixed-effects results in the Appendix (3.5) under Table 3.3. The random-effects model corroborates the results of the fixed-effects approach, as the simultaneous SFR variable is insignificant while a change by one in all lagged SFR variables account for a more than 3% increase in the number of SuTP in a province. While the results are in line with the fixed-effects model, we opt for the fixed-effects approach based on the results of the Hausman test. Then, we replace the SFR variable with SF in Arabic letters and SF in Turkish letters as two separate variables, both to test our approach of using search frequency ratio as the explanatory variable and to disentangle their relationship with the dependent variable of number of SuTP. The results are reported in the Appendix (3.5) under Table 3.4. The results show that similar to the results of our initial models with controls, the simultaneous online SF variable is neither for the SF in Turkish letters nor for the SF in Arabic letters significant. All lagged versions of SF in Arabic letters are significant and positively correlated with the number of SuTP. One-month lag appears insignificant for SF in Turkish letters, but both one-week and six-week lags are positively and significantly associated with the dependent variable.

Next, we replicated the model using the same queries and time period for the online SF variable, changing the online search location from Turkey to Syria. As the online search frequencies for Turkish province names in Syria includes considerable amount of zero values, we used the SF in Arabic letters and SF in Turkish letters variables separately instead of calculating SFR variables. The results shown in the Appendix (3.5), in Table 3.5 demonstrate that online searches for Turkish province names in Turkish letters is a sign of intention to move to that province. An increase by one in all three lagged SF in Turkish letters variables are correlated with around 9% increase in the number of SuTP in a given province in Turkey. As the online search location is Syria, we must bear in mind that the online search may be conducted not by the Syrian citizens who are about to move to the said province in Turkey but instead the relatives of SuTP in Turkey who are planning to that province as well.

Last, we replaced the SF in Turkish letters with SF in English letters, i.e. removed the Turkish special characters from the online search query and repeated the analysis. The results for Turkey and Syria as the location of online searches are reported in the Appendix (3.5), in tables 3.6 and 3.7 respectively. Comparing the results in Table 3.6 with the initial results, one can see that the two are quite similar. A possible explanation for this may be that Google algorithm is sensitive to the spelling of search queries with and without Turkish special characters in Turkey. Considering that Google Trends data is produced on a different and representative sample of all searches at every request, controlling for no Turkish characters and obtaining similar results also serves as a control for the same data taken from a different sample. Use of English characters, however, creates a difference for the analysis in Syria and the association with the number of SuTP becomes stronger (Table 3.7).

3.4 Conclusion

This paper aims to shed some light into the relationship between online search intensity and people's mobility from one place to another, by examining the case of SuTP in Turkey. Assuming that Syrians would seek information online in a foreign country they are not familiar with and building on the previous research indicating high smart phone & internet usage by Syrian refugees (Ulutürk et al., 2019); we tested this potential relationship between Google Trends data for province names in Turkey and number of SuTP across provinces over a period of four years. The results demonstrate that there is a positive and significant association between the online search frequency for province names in Turkey and number of SuTP in the same provinces. However, this association is observed not for the same week as the SuTP statistics are published, but for one week, one month and six weeks prior. The association is observed stronger for the SFR with a six weeks lag in respect to the week the statistics are published, followed by the one month lag. This result is coherent with the time needed for preparation before travel and necessary bureaucratic steps. The online search frequency for Turkish province names in Syria using Latin letters, both with and without Turkish special characters, is also positively and significantly correlated with the number of SuTP across provinces, supporting the results found for Turkey as the online search location. The association is robust according to the all checks and controls applied and even though small in size, it corroborates with the pattern of small but significant correlation found in previous studies (Choi and Varian, 2009).

These results suggests that in the case of SuTP in Turkey, the online search patterns indicate an intention to move but the behaviour resulting out of this intention may be observed in the official registries after a one month to six weeks interval. In other words, the results of this study suggests that *forecasting* the number of SuTP in provinces is possible, based on the association observed with the online search data patterns. As we cannot find a significant relationship between the online search frequency for province names and number of SuTP across provinces for the week the statistics are updated, *nowcasting* cannot be used as a method to the observe a change in SuTP statistics. However, as we find a significant relationship between SFR and SuTP statistics with a time lag, it is plausible to argue that *nowcasting* can be a method to identify the mobility decisions, in this case, of Syrians in Turkey.

Although the findings of this study suggest a significant and positive correlation between search frequency for province names and the number of SuTP across provinces, policy implications of these findings are harder to pin down. Based on the positive and significant association this study proposes, future research may look for further correlations considering an interaction between the search queries made for the name of provinces, search queries made for standard phrases such as looking for jobs or apartments for rent in Arabic and in Turkish, in Syria and in Turkey and the geolocation of these searches (available at sub-regional level on Google Trends, that is province level for Turkey). Such studies based on the effect of interactions may exploit the difference of both the alphabet and language. Furthermore, the exploitation of the difference in alphabet can also be implemented in other non-Arabic speaking and refugee hosting countries, to both compare the results and obtain better insight on Syrian refugees mobility patterns. Correlations between such search queries and actual number of SuTP, if proven, may help policy-makers at urban level to prepare for incoming refugees as well as national and international level policy-makers to better plan the social assistance programs.

3.5 Appendix - Chapter 3

Table 3.2: Descriptive statistics

Variable		Mean	St. Dev.	Min	Max	Observations
SF	overall	57.08	16.38	8	100	N = 9,882
Turkish let.	between		13.93	14.34	79.62	n = 81
in Turkey	within		8.75	33.95	139.37	T = 122
SF	overall	7.46	18.59	0	100	N = 9,882
Turkish let.	between		7.80	0	37.96	n = 81
in Syria	within		16.90	-30.50	105.40	T = 122
SF	overall	55.25	16.76	7	100	N = 9,882
English let.	between		14.20	14.12	80.21	n = 81
in Turkey	within		9.04	28.514	137.26	T = 122
SF	overall	8.26	19.08	0	100	N = 9,882
English let.	between		10.07	0	35.25	n = 81
in Syria	within		16.25	-26.99	105.21	T = 122
SF	overall	19.74	22.57	0	100	N = 9,882
Arabic letters	between		10.94	0	47.64	n = 81
in Turkey	within		19.78	-17.05	115.63	T = 122
SF	overall	6.70	16.48	0	100	N = 9,882
Arabic letters	between		11.10	0	47.10	n = 81
in Syria	within		12.25	-37.12	105.88	T = 122
	overall	41,032.93	100,257.60	14	564,189	N = 9,882
#SuTP	between		100,148.70	34.17	511,092	n = 81
	within		12,026.84	-112,834.10	116,794.60	T = 122
	overall	8.20	2.40	2.64	13.24	N = 9,882
ln(#SuTP)	between		2.39	3.49	13.14	n = 81
	within		0.31	6.18	9.66	T = 122

Table 3.2 Descriptive Statistics cont.

Variable		**Mean**	**St. Dev.**	**Min**	**Max**	**Observations**
	overall	13.25	0.95	11.27	16.53	N = 9,882
ln(Population)	between		0.96	11.33	16.52	n = 81
	within		0.02	13.16	13.36	T = 122
	overall	0.35	0.45	0	7.08	N = 9,882
SFR	between		0.20	0	1.11	n = 81
	within		0.41	-0.53	6.97	T = 122
	overall	0.34	0.46	0	7.08	N = 9,882
$\mathbf{SFR_{tj\text{-}1}}$	between		0.20	0	1.08	n = 81
	within		0.42	-0.37	6.94	T = 122
	overall	0.34	0.45	0	7.08	N = 9,882
$\mathbf{SFR_{tj\text{-}4}}$	between		0.19	0	1.08	n = 81
	within		0.41	-0.32	6.92	T = 122
	overall	0.34	0.43	0	7.08	N = 9,882
$\mathbf{SFR_{tj\text{-}6}}$	between		0.20	0	1.07	n = 81
	within		0.39	-0.35	6.97	T = 122

Table 3.3: Comparison of the Fixed-effects and Random-effects Results

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	ln(#SuTP)	ln(#SuTP)	ln(#SuTP)	ln(#SuTP)	ln(#SuTP)	ln(#SuTP)	ln(#SuTP)	ln(#SuTP)
$Arabic/Turkish\ SFR_{tj}$	0.0928***	0.0932***	0.0673***	0.0676***	0.0245**	0.0361***	0.00241	0.0126
	(0.0236)	(0.0236)	(0.0151)	(0.0152)	(0.0111)	(0.0115)	(0.0200)	(0.0194)
$Arabic/Turkish\ SFR_{tj-1}$			0.0607***	0.0610***	0.0206*	0.0315***	0.0194*	0.0302**
			(0.0164)	(0.0164)	(0.0109)	(0.0117)	(0.0110)	(0.0119)
$Arabic/Turkish\ SFR_{tj-4}$			0.0658***	0.0661***	0.0231**	0.0347***	0.0231**	0.0349***
			(0.0137)	(0.0137)	(0.0104)	(0.0106)	(0.0103)	(0.0106)
$Arabic/Turkish\ SFR_{tj-6}$			0.0749***	0.0752***	0.0261**	0.0393***	0.0256**	0.0390***
			(0.0166)	(0.0166)	(0.0127)	(0.0132)	(0.0126)	(0.0133)
ln(Population)					6.406***	4.626***	6.286***	4.436***
					(1.207)	(0.668)	(1.194)	(0.633)
No SF in Turkish Letters							-0.0587***	-0.0738***
(SYR)							(0.0191)	(0.0187)
No SF in Arabic Letters							-0.00495	-0.00479
(SYR)							(0.0142)	(0.0142)
No SF in Arabic Letters							-0.0363	-0.0388
(TR)							(0.0345)	(0.0343)
Constant	8.168***	8.168***	8.108***	8.108***	-76.71***	-53.15***	-75.04***	-50.54***
	(0.00813)	(0.263)	(0.0196)	(0.261)	(16.00)	(8.930)	(15.81)	(8.452)
Observations	9,882	9,882	9,882	9,882	9,882	9,882	9,882	9,882
R-squared (within)	0.014		0.041		0.207		0.211	
Overall R-squared	0.0881	0.0881	0.182	0.182	0.563	0.564	0.564	0.566
Number of Provinces	81	81	81	81	81	81	81	81
FE	YES	NO	YES	NO	YES	NO	YES	NO
RE	NO	YES	NO	YES	NO	YES	NO	YES

Robust standard errors in parentheses *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 3.4: Using SF variables separately instead of SFR

VARIABLES	(1) ln(#SuTP)	(2) ln(#SuTP)	(3) ln(#SuTP)	(4) ln(#SuTP)	(5) ln(#SuTP)	(6) ln(#SuTP)	(7) ln(#SuTP)	(8) ln(#SuTP)
Arabic Letters SF_{tj}	0.00188***	0.00189***	0.000939***	0.000945***	0.000427**	0.000573***	2.81e-05	0.000195
	(0.000347)	(0.000347)	(0.000204)	(0.000203)	(0.000193)	(0.000186)	(0.000530)	(0.000519)
Arabic Letters SF_{tj-1}			0.000947***	0.000953***	0.000457**	0.000598***	0.000463**	0.000602***
			(0.000212)	(0.000212)	(0.000204)	(0.000193)	(0.000210)	(0.000202)
Arabic Letters SF_{tj-4}			0.000925***	0.000931***	0.000391*	0.000544***	0.000410**	0.000564***
			(0.000212)	(0.000212)	(0.000203)	(0.000195)	(0.000199)	(0.000192)
Arabic Letters SF_{tj-6}			0.000974***	0.000981***	0.000390*	0.000557***	0.000391*	0.000561***
			(0.000238)	(0.000238)	(0.000211)	(0.000210)	(0.000210)	(0.000211)
Turkish Letters SF_{tj}	0.00685***	0.00685***	0.00200***	0.00200***	0.000879	0.00118*	0.000795	0.00108
	(0.00148)	(0.00148)	(0.000689)	(0.000688)	(0.000673)	(0.000666)	(0.000743)	(0.000735)
Turkish Letters SF_{tj-1}			0.00215***	0.00215***	0.00110**	0.00139***	0.00112**	0.00143***
			(0.000538)	(0.000537)	(0.000469)	(0.000479)	(0.000471)	(0.000484)
Turkish Letters SF_{tj-4}			0.00248***	0.00247***	0.000927	0.00136**	0.000905	0.00135**
			(0.000670)	(0.000669)	(0.000611)	(0.000618)	(0.000613)	(0.000622)
Turkish Letters SF_{tj-6}			0.00415***	0.00415***	0.00234***	0.00285***	0.00234***	0.00286***
			(0.000837)	(0.000836)	(0.000747)	(0.000771)	(0.000744)	(0.000770)
ln(Population)					5.570***	3.961***	5.501***	3.824***
					(1.142)	(0.588)	(1.134)	(0.559)
No SF in Turkish Let.							-0.0501***	-0.0592***
(SYR)							(0.0183)	(0.0178)

Table 3.4 cont.

VARIABLES	(1) ln(#SuTP)	(2) ln(#SuTP)	(3) ln(#SuTP)	(4) ln(#SuTP)	(5) ln(#SuTP)	(6) ln(#SuTP)	(7) ln(#SuTP)	(8) ln(#SuTP)
No SF in Arabic Letters (SYR)							-0.00149	-0.000369
							(0.0126)	(0.0121)
No SF in Arabic Letters (TR)							-0.0280	-0.0268
							(0.0385)	(0.0383)
Constant	7.772***	7.771***	7.512***	7.512***	-65.94***	-44.71***	-64.96***	-42.82***
	(0.0876)	(0.282)	(0.143)	(0.312)	(15.11)	(7.852)	(15.01)	(7.467)
Observations	9,882	9,882	9,882	9,882	9,882	9,882	9,882	9,882
R-squared (within)	0.062		0.117		0.223		0.226	
Overall R-squared	0.0419	0.0420	0.0578	0.0581	0.561	0.560	0.562	0.562
Number of Provinces	81	81	81	81	81	81	81	81
FE	YES	NO	YES	NO	YES	NO	YES	NO
RE	NO	YES	NO	YES	NO	YES	NO	YES

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 3.5: Replication for Syria as the location of online searches

VARIABLES	(1) ln(#SuTP)	(2) ln(#SuTP)	(3) ln(#SuTP)	(4) ln(#SuTP)	(5) ln(#SuTP)	(6) ln(#SuTP)	(7) ln(#SuTP)	(8) ln(#SuTP)
Arabic Letters SF_{tj}	0.000256	0.000273	0.000205	0.000218	7.85e-05	9.54e-05	-5.63e-05	-3.83e-05
	(0.000382)	(0.000382)	(0.000320)	(0.000319)	(0.000288)	(0.000288)	(0.000375)	(0.000367)
Arabic Letters SF_{tj-1}			-0.000207	-0.000193	-0.000431*	-0.000389	-0.000435*	-0.000392
			(0.000298)	(0.000297)	(0.000257)	(0.000258)	(0.000258)	(0.000263)
Arabic Letters SF_{tj-4}			0.000145	0.000158	5.28e-05	6.11e-05	5.40e-05	6.29e-05
			(0.000243)	(0.000242)	(0.000208)	(0.000209)	(0.000216)	(0.000217)
Arabic Letters SF_{tj-6}			-0.000325	-0.000311	-0.000404*	-0.000398*	-0.000411*	-0.000408*
			(0.000277)	(0.000276)	(0.000230)	(0.000236)	(0.000230)	(0.000238)
Turkish Letters SF_{tj}	0.00224***	0.00225***	0.00194***	0.00195***	0.00101***	0.00125***	0.000533	0.000610*
	(0.000351)	(0.000350)	(0.000314)	(0.000313)	(0.000302)	(0.000291)	(0.000373)	(0.000351)
Turkish Letters SF_{tj-1}			0.00153***	0.00154***	0.000655**	0.000878***	0.000643**	0.000890***
			(0.000283)	(0.000283)	(0.000256)	(0.000251)	(0.000258)	(0.000255)
Turkish Letters SF_{tj-4}			0.00162***	0.00162***	0.000673***	0.000914***	0.000663***	0.000931***
			(0.000249)	(0.000248)	(0.000246)	(0.000230)	(0.000243)	(0.000228)
Turkish Letters SF_{tj-6}			0.00156***	0.00156***	0.000680***	0.000902***	0.000660**	0.000904***
			(0.000291)	(0.000291)	(0.000256)	(0.000255)	(0.000253)	(0.000253)

Table 3.5 cont.

VARIABLES	(1) ln(#SuTP)	(2) ln(#SuTP)	(3) ln(#SuTP)	(4) ln(#SuTP)	(5) ln(#SuTP)	(6) ln(#SuTP)	(7) ln(#SuTP)	(8) ln(#SuTP)
ln(Population)					6.405***	4.740***	6.313***	4.410***
					(1.213)	(0.706)	(1.198)	(0.630)
No SF in Turkish Letters (SYR)							-0.0311	-0.0435*
							(0.0241)	(0.0229)
No SF in Arabic Letters (SYR)							-0.00888	-0.00916
							(0.0168)	(0.0159)
No SF in Arabic Letters (TR)							-0.0397*	-0.0520**
							(0.0229)	(0.0235)
Constant	8.181***	8.181***	8.152***	8.152***	-76.68***	-54.63***	-75.41***	-50.18***
	(0.00458)	(0.265)	(0.0132)	(0.265)	(16.07)	(9.436)	(15.87)	(8.404)
Observations	9,882	9,882	9,882	9,882	9,882	9,882	9,882	9,882
R-squared (within)	0.015		0.039		0.211		0.213	
Overall R-squared	0.0752	0.0760	0.134	0.136	0.563	0.564	0.564	0.565
Number of Provinces	81	81	81	81	81	81	81	81
FE	YES	NO	YES	NO	YES	NO	YES	NO
RE	NO	YES	NO	YES	NO	YES	NO	YES

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 3.6: No Turkish characters in search queries, location Turkey

VARIABLES	(1) ln(#SuTP)	(2) ln(#SuTP)	(3) ln(#SuTP)	(4) ln(#SuTP)	(5) ln(#SuTP)	(6) ln(#SuTP)	(7) ln(#SuTP)	(8) ln(#SuTP)
$Arabic/English\ SFR_{tj}$	0.0983***	0.0986***	0.0688***	0.0691***	0.0272**	0.0386***	0.0123	0.0238
	(0.0229)	(0.0229)	(0.0140)	(0.0140)	(0.0105)	(0.0108)	(0.0193)	(0.0190)
$Arabic/Engl.\ SFR_{tj-1}$			0.0659***	0.0661***	0.0260**	0.0369***	0.0248**	0.0357***
			(0.0154)	(0.0154)	(0.0109)	(0.0114)	(0.0108)	(0.0116)
$Arabic/Engl.\ SFR_{tj-4}$			0.0690***	0.0692***	0.0267***	0.0383***	0.0266***	0.0385***
			(0.0124)	(0.0124)	(0.00994)	(0.00990)	(0.00984)	(0.00992)
$Arabic/Engl.\ SFR_{tj-6}$			0.0745***	0.0748***	0.0287**	0.0412***	0.0284**	0.0412***
			(0.0152)	(0.0152)	(0.0116)	(0.0122)	(0.0116)	(0.0123)
ln(Population)					6.281***	4.531***	6.172***	4.337***
					(1.199)	(0.657)	(1.186)	(0.620)
No SF in English Lett. (SYR)							-0.0577***	-0.0723***
							(0.0191)	(0.0189)
No SF in Arabic Letters (SYR)							-0.00432	-0.00414
							(0.0143)	(0.0142)
No SF in Arabic Letters (TR)							-0.0266	-0.0267
							(0.0347)	(0.0344)
Constant	8.164***	8.164***	8.100***	8.100***	-75.07***	-51.89***	-73.55***	-49.24***
	(0.00832)	(0.264)	(0.0188)	(0.262)	(15.89)	(8.792)	(15.72)	(8.282)
Observations	9,882	9,882	9,882	9,882	9,882	9,882	9,882	9,882
R-squared	0.019		0.053		0.210		0.213	
Overall R2	0.0801	0.0801	0.166	0.166	0.563	0.564	0.564	0.566
Number of Provinces	81	81	81	81	81	81	81	81
FE	YES	NO	YES	NO	YES	NO	YES	NO
RE	NO	YES	NO	YES	NO	YES	NO	YES

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 3.7: No Turkish characters in search queries, location Syria

VARIABLES	(1) ln(#SuTP)	(2) ln(#SuTP)	(3) ln(#SuTP)	(4) ln(#SuTP)	(5) ln(#SuTP)	(6) ln(#SuTP)	(7) ln(#SuTP)	(8) ln(#SuTP)
Arabic Letters SF_{tj}	0.000161	0.000178	0.000119	0.000133	2.19e-05	3.15e-05	-0.000154	-0.000150
	(0.000375)	(0.000374)	(0.000301)	(0.000300)	(0.000275)	(0.000273)	(0.000343)	(0.000326)
Arabic Letters SF_{tj-1}			-0.000230	-0.000216	-0.000450*	-0.000411	-0.000453*	-0.000412
			(0.000285)	(0.000285)	(0.000252)	(0.000252)	(0.000260)	(0.000262)
Arabic Letters SF_{tj-4}			2.75e-05	4.10e-05	-1.09e-05	-1.55e-05	-1.13e-05	-1.81e-05
			(0.000238)	(0.000238)	(0.000201)	(0.000202)	(0.000207)	(0.000209)
Arabic Letters SF_{tj-6}			-0.000356	-0.000342	-0.000425*	-0.000422*	-0.000433*	-0.000434*
			(0.000261)	(0.000260)	(0.000218)	(0.000222)	(0.000221)	(0.000227)
English Letters SF_{tj}	0.00242***	0.00243***	0.00204***	0.00205***	0.00118***	0.00138***	0.000892***	0.000986***
	(0.000384)	(0.000383)	(0.000308)	(0.000307)	(0.000286)	(0.000278)	(0.000244)	(0.000243)
English Letters SF_{tj-1}			0.00176***	0.00177***	0.000958***	0.00115***	0.000944***	0.00116***
			(0.000350)	(0.000349)	(0.000322)	(0.000320)	(0.000322)	(0.000323)
English Letters SF_{tj-4}			0.00176***	0.00177***	0.000822***	0.00105***	0.000815***	0.00107***
			(0.000302)	(0.000301)	(0.000278)	(0.000269)	(0.000274)	(0.000266)
English Letters SF_{tj-6}			0.00176***	0.00176***	0.000960***	0.00115***	0.000951***	0.00117***
			(0.000318)	(0.000318)	(0.000292)	(0.000290)	(0.000291)	(0.000291)

Table 3.7 cont.

VARIABLES	(1) ln(#SuTP)	(2) ln(#SuTP)	(3) ln(#SuTP)	(4) ln(#SuTP)	(5) ln(#SuTP)	(6) ln(#SuTP)	(7) ln(#SuTP)	(8) ln(#SuTP)
ln(Population)					6.357***	4.756***	6.246***	4.368***
					(1.193)	(0.710)	(1.179)	(0.625)
No SF in English Letters (SYR)							-0.0264*	-0.0396***
							(0.0156)	(0.0147)
No SF in Arabic Letters (SYR)							-0.0120	-0.0130
							(0.0163)	(0.0151)
No SF in Arabic Letters (TR)							-0.0404*	-0.0531**
							(0.0229)	(0.0234)
Constant	8.179***	8.179***	8.144***	8.143***	-76.06***	-54.85***	-74.54***	-49.64***
	(0.00519)	(0.265)	(0.0149)	(0.264)	(15.80)	(9.491)	(15.63)	(8.336)
Observations	9,882	9,882	9,882	9,882	9,882	9,882	9,882	9,882
R-squared (within)	0.016		0.044		0.215		0.218	
Overall R-squared	0.109	0.109	0.188	0.190	0.563	0.564	0.564	0.566
Number of Provinces	81	81	81	81	81	81	81	81
FE	YES	NO	YES	NO	YES	NO	YES	NO
RE	NO	YES	NO	YES	NO	YES	NO	YES

Robust standard errors in parentheses

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Chapter 4

Brexit and academic migration involving the UK: Evidence from affiliation addresses in Scopus publications

Ebru Sanliturk [1,2], Samin Aref [2], Emilio Zagheni[2], Francesco C. Billari [1,3]

1. Bocconi University
2. Max Planck Institute for Demographic Research
3. Carlo F. Dondena Centre for Research on Social Dynamics and Public Policy

Abstract

The literature on migration argues that certain factors and conditions generate a migration momentum. Structural factors such as labour demand, migration networks as well as sudden discontinuities such as wars, famines, epidemics, technological transformations and policy change are, in that line, reasons for a change in migration flows. Migration policies address these factors by introducing regulations and opportunities to attract the labour force and highly skilled individuals, to meet the countries' requirements. The United Kingdom's decision to exit from the European Union (EU) (hereafter Brexit) in 2016 constitutes a political discontinuity, and whether Brexit will be a blip in history without long-term consequences or alter the migration trends, especially those of high-skilled people, is a highly debated question. This study assesses the initial effects of Brexit on the mobility of scholars, as a subgroup of high-skilled migrants, to and from the United Kingdom (UK). We use an exhaustive database of the affiliation addresses of researchers from all Scopus-indexed publications over the past two decades. These data enable us to assess changes in the international migration of researchers before and after the Brexit referendum. Although it is too early to talk about a brain drain following the Brexit, considering that Brexit came into effect in 2020, our scholars began to change after 2016. Among the UK-resident scholars, the odds of leaving the UK after Brexit is 35% higher if their academic origin (country of first publication) is an EU country, as opposed to the 21% lower odds of leaving the UK observed for the same group of scholars without the Brexit condition. Furthermore, the odds of returning to the UK is 18% higher for scholars with UK academic origin after Brexit, in stark contrast to the pre-Brexit situation, increasing their share among overall incoming researchers. Our analysis of Scopus data points to a compositional change in terms of academic origin, as the first impact of Brexit on UK academic environment.

4.1 Introduction

International circulation of scholars is fundamental to foster scientific knowledge, especially in its most innovative forms (Agrawal et al., 2014; Fernández-Zubieta et al., 2016). For instance, nearly half of the world's most cited physicists reside outside their country of birth (Hunter et al., 2009). The international mobility of scientists is therefore a special case of high-skilled migration which rightly commands attention from researchers and policy-makers alike (Czaika, 2018). For these reasons, it is paramount to understand the dynamics of in- and out-flows of scholars between countries, and the underlying determinants of the international mobility of scientists.

Within the international migration literature, academic migration that has been studied in the framework of the brain drain and brain gain relationship, can be aptly framed through the concept of brain circulation (Saxenian, 2005). While many reasons influence the decisions of scholars to move (Azoulay et al., 2017), a key determinant is the policy environment. More specifically, policy changes may substantially affect the decisions of scientists to migrate internationally (Arrieta et al., 2017; Scellato et al., 2015; Franzoni et al., 2015) and in turn impact the scientific and technological development of the countries involved (Mahroum, 2005; Moser et al., 2014).

The UK decided to withdraw from the EU as a result of the Referendum held on June 23rd, 2016, and *Brexit* became official on January 31st, 2020. Despite the long-lasting fears about losing scientists to other countries (Irvine et al., 1985; Martin et al., 1987; Martin, 1994), the UK is one of the leading nations in scientific research in the world. In 2019, the UK had the highest share in the globally top 10% of high-quality scientific papers among G20 countries (Adams et al., 2019). As a (now former) member of the EU, the UK possessed a competitive advantage, due to its scientific performance, in terms of benefiting from research funding provided by the EU. The UK has benefited from the EU also in terms of human capital; the share of European nationals among overall British academic staff was estimated as 16% and among the technology industry in London as 20% by 2018 (Main, 2018). The ties British science and technology has established with the EU are the reason why the scientists fear from a backlash due to Brexit (Golding, 2017).

This study aims to analyse brain circulation patterns in the UK, and, in particular, the potential effects of the decision to withdraw from the EU, Brexit. The contributions of this study are twofold. First, we contribute to the high-skilled migration literature by providing timely evidence for the changing circulation patterns of highly productive researchers in the UK academic environment by re-purposing bibliometric data using machine learning methods. Second, we shed light to the discussions on how Brexit would affect the scientific environment in the UK, by introducing decrease in diversity in terms of academic background as a potential result of Brexit in contrast to an immediate brain drain.

4.2 Literature Review

4.2.1 High-skilled Migration

Skill level of migrants is long considered as an important aspect of migration, as migrants with different skill levels are often influenced by different push and pull effects, have different implications on the economy and development of both source and destination countries and require different policies. In that respect, high-skilled migration refers to the migration processes where the migrants have specialized formal education and/or advanced qualifications, such as scientists, doctors, entrepreneurs etc (Nathan, 2014). High-skilled migration vastly addressed in the literature both with a macro perspective, investigating its implications (Regets, 2001; Ouaked, 2002; Chiswick, 2005; Leipziger, 2008) and policy requirements in source and destination countries (Czaika and Parsons, 2017; Hercog and Sandoz, 2018), and with a micro perspective examining the decision-making processes of high-skilled migrants (Sbalchiero and Tuzzi, 2017).

High-skilled migration has initially been considered in the context of developing and developed countries, in terms of the *brain drain* from the former and *brain gain* of the latter. The common perspective in this approach is that the flow of high-skilled migrants from developing countries to the developed countries creates a continuous negative effect for the developing countries due to the loss of the human capital (Bhagwati and Hamada, 1974; Miyagiwa, 1991; Haque and Kim, 1995).

More recently, empirical studies addressed the brain drain - brain gain dichotomy critically and examined the potential benefits of high-skilled emigrants for developing countries (Nathan, 2014). We can name four four ways, where emigration of high-skilled citizens can have positive outcomes for the source country as examples. The first is the benefit of remittances for the economy of the developing countries, which is argued to be higher in the case of high-skilled, higher educated migrants as earn higher income in comparison to the low-skilled labour migrants (Bollard et al., 2011). Studies also show that host countries' immigration policies may influence the potential upper-hand of high-skilled migrants in terms of remittances positively or negatively (Docquier et al., 2012). Second, in as demonstrated in the influential studies, the migration expectation may create incentives for an higher investment in education in developing countries and an overall increase for human capital (Stark et al., 1998; Beine et al., 2001). Third, high-skilled migrants living in developed countries could contribute to the development of their home countries through positive network externalities, their connections, transfer of resources & knowledge as well as diaspora networks (Docquier and Rapoport, 2012). The case of high-skilled migrants' diaspora structures and development in Colombia is a prominent example of this theory (Meyer et al., 1997). Last but not least, the returning high-skilled migrants may bring their knowledge and connections to foster economic and technological development at home (Chacko, 2007).

Using the case of the Silicon Valley in the U.S. and China & India, Saxenian ()saxenian2005brain theorizes that the high-skilled migrants' connections to the low-cost home country labour market as well as the returning high-skilled migrants foster the specialization of advanced technologies in developing countries. As this pattern becomes more common in high-technology fields, it replaces the brain drain-brain gain dichotomy with the concept of *brain circulation*, where the mobility and migration of high-skilled people work and benefit both ways (Saxenian, 2005). In a similar understanding of constant movement and mutual benefit, the concept of brain circulation is used also for the high-skilled migration between developed countries, especially in the context of European Union, owing to the lack of mobility restriction among member countries (Ouaked, 2002). Considering the policy intervention this chapter uses as a case study is the United Kingdom's withdrawal from the European Union and bearing in mind the close ties that the United Kingdom has with Commonwealth countries and other English-speaking countries, we use the concept of brain circulation to analyse the scientific migration to and from the United Kingdom.

4.2.2 The United Kingdom as a Brain Circulation Hub

The notion of brain circulation has been present in scientific debate within the UK for a long time. In fact, the term brain drain has been coined in this very context. During the early-1960s the Royal Society published a report on the increase in emigration of scientists and engineers from the UK to the USA or Canada, referring this situation as "a drain of scientists and drain of talent" (Oldfield et al., 1963). The drain of scientists and talent out of the UK was later defined as *brain drain* (Johnson, 1965).

Concerns about the brain drain have lessened during the 1970s, as both the British policymakers started to consider brain drain as an inevitable part of globalization, and the US became less appealing for scientists due to the Vietnam War (Godwin et al., 2009). However, in the 1980s, fears of a decline in British science reappeared among academics. The performance of British science in STEM fields, measured by the share of publications and citations in the world, has been observed to decrease by %10 and %15 respectively between 1973-1982, while the sharpest decline occurred in Physics, Engineering, and Technology fields with over %20 (Irvine et al., 1985, p.588). In reaction to these concerns, the initiative *Save British Science* was launched in 1986, calling for the government to take action and support research as "opportunities are missed, scientists emigrate and whole areas of research is in jeopardy" (Noble, 2016). During the early 1990s, research showed that the scientific performance of Britain grew in some areas, but overall the relative decline was persisting (Martin, 1994).

Although the general impression about the performance of British science has been a rather pessimistic one since the early 1960s, one should refrain from over-emphasising the lack of scientific investment and scientists' emigration as the underlying reason. It should be noted that the apparent decline in scientific performance is also partly due to the increased ability of scientists worldwide to publish in English language, which mitigated the native English-speaker bias to a certain degree. Furthermore, starting from late 1960, the emigration of scientists from the UK to the US and Canada has been compensated by the immigration of scientists from developing countries (and/or Commonwealth countries) to the UK (Anonymous, 1967/04; Watanabe, 1969; Godwin et al., 2009).

Hence, despite the concerns about a decline in scientific performance, the UK has continued to be one of the top countries in science. According to *Campaign for Science and Engineering* (CaSE), the advocacy group and successor of the *Save British Science* initiative, the UK is a science super-power that has substantially benefited from European Union's research funding (Main, 2018). Recent figures show that the UK has received a total of 8.8 billion Euros for R&D and innovation between 2007-2013 (Frenk et al., 2015) (7^{th} Framework Program and Structural Funds) and a further 6.43 billion Euros between 2014-2020 (Horizon 2020) (European Commission, 2020). The UK has benefited from the EU also in terms of human capital; the share of European nationals among British academic staff was estimated as %16 and among the technology industry in London as %20 by 2018 (Main, 2018). The ties British science and technology has established with the European Union (EU) are the reason why the scientists fear from a backlash due to Brexit (Golding, 2017).

4.2.3 Big Bibliometric Data and Scientific Migration

The early use of bibliometric data, limited in terms of the volume of data used, focused more on using citation count as the unit of measure to assess scientific impact, scientific progress (Martin and Irvine, 1983), and institutional research performance (Moed et al., 1985). The assessment of scientific performance through bibliometric data influenced not only scholars but also policy-makers during the 1990s, especially under the New Public Management framework (Mingers and Leydesdorff, 2015). Over the past decades, the volume of data used for bibliometric analyses has expanded in time, and beyond the country- or institutional-level, creating what one could term big bibliometric data. As the literature on measuring scientific performance using bibliometric data continued to grow (Sugimoto and Larivière, 2018), big bibliometric data paved a new ground of study for migration research (Alburez-Gutierrez et al., 2019).

Migration studies using bibliometric data rely on the information on movements of researchers. Following the network-based approach in high-skilled migration (Meyer, 2001) and scientific migration (Ackers, 2005), the use of bibliometric data to study researchers' migration and mobility started to receive some attention (Laudel, 2003). The feasibility of the method in examining the scientific migration and mobility patterns was demonstrated first in the cases of a selection of countries (Moed et al., 2013; Moed and Halevi, 2014). More recently the literature on scientific migration using bibliometric data started to grow with studies addressing co-affiliation and collaboration networks (Sugimoto et al., 2016; Aref et al., 2018), identification of migration and mobility events (Robinson-García et al., 2019) and mobility patterns of highly mobile researchers (Aref et al., 2019).

The use of bibliometric data also presented a tool to study certain demographic characteristics of researchers. Gender disparities and their influence on scientific performance (Larivière et al., 2013), academic age of researchers (Nane et al., 2017), and the impact of academic age on international mobility (Sugimoto et al., 2017) are known as the prominent studies in that area.

4.3 Data and Methods

The data we use in this study is obtained from Scopus, a database containing detailed meta-data on scientific publications indexed by Elsevier's abstract and citation database. The database includes detailed meta-data for each publication, that includes key information for the purposes of this study such as; individual author ID, publication year, and affiliation country per publication and author ID, and All Science Journal Classification (ASJC) code for field per publication venue. To obtain the raw bibliometric data, we query an extract of all Scopus data from a relational database using SQL. The query involves two steps: (1) obtaining author IDs of all authors who match the criteria relevant to this study, and (2) obtaining all authorship records for the list of author IDs produced in the previous step. We have two selection criteria for any researcher to be included in our study: (1) to have more than one country of affiliation in their academic career (international mobility), and (2) to have at least one publication with a UK affiliation. The raw data is then pre-processed to be used in empirical analyses, which most importantly address the challenges of missing values for the country variable and author name ambiguity.

We build on previous research on bibliometric data to define academic migration. In what follows, the country of academic origin is defined as the country of first publication. The academic origin is not considered as a proxy for the nationality of a scholar, but as their origin in an academic sense. The country of academic origin is therefore the one that has invested in their academic development, regardless of the nationality of the scholar (Robinson-García et al., 2016, 2019; Aref et al., 2019; Subbotin and Aref, 2020). Similarly, academic age is measured using the year of the first publication. Academic migration across countries is defined as a change in the country of affiliation in different years for a given scholar. The number of these changes are used to estimate academic migration flows.

We focus on yearly migratory movements rather than on short-term mobility for two reasons. First, the literature on migration helps us to understand these flows through concepts such as high-skilled migration, elite migration, and brain-drain (Laudel, 2005). Second, considering the frequency of short-term international moves in academia, the academic migration argument allows us to differentiate cross-country movements from long-lasting migratory moves, and therefore examine the effect of country-specific policies that have a potential impact on a country's productivity. Furthermore, we acknowledge the presence of short-term cross-country mobility for scientific collaborations in the academia, and therefore apply a mobility taxonomy to identify migration events and distinguish them from cross-country affiliations.

In contrast to the mobility taxonomy proposed in (Robinson-García et al., 2019) this study considers the modal (the most frequent) country of affiliation for each year among all authorship records of an author as an input to determine residence. Migration events are therefore defined a changes of residence. The logic behind this approach is to reduce the false positive error that arise in considering small changes in affiliation as migratory behaviour. Note that, sometimes, changes in affiliation could be due to other reasons such as the requirements of a foreign funding body, while no geographical mobility is taking place.

More precisely, we define a migration event when two requirements are satisfied: (1) a new country appears as the modal country of affiliation of a scholar and (2) the previous modal country of affiliation disappears as a mode (or from the list of modes in case of multiple modes). In the rare cases when a researcher's academic career, as observed in Scopus database, begins with multiple mode countries of affiliations in the first year of publication, we randomly assign the researcher's residence to one of the countries that appears as the mode. This way, we ensure that the researchers' careers begin with a single country, and avoid possible bias through randomization. In other cases, where a researcher has multiple mode country affiliations in a year, we apply a taxonomy which is outlined in Table 4.1 below.

The data we use in this study is obtained from Scopus, a database containing detailed meta-data on scientific publications indexed by Elsevier's abstract and citation database. The database includes detailed meta-data for each publication but the key information we obtain by querying the database are shown in Table 4.3 in the Appendix–Chapter 4, with a brief description on what they indicate for the purposes of this study. Publication venues (journals, conference proceedings, etc.) are tagged based on All Science Journal Classification (hereafter ASJC) codes into four general categories: life sciences, social sciences, physical sciences, and health sciences.

Last, we create a gender indicator, inferred by the first names of researchers using genderizeR package on R (Wais, 2006). The use of gender estimation algorithms in big bibliometric data analysis is relatively new and as the initial development of such algorithms had marketing purposes, Asian and African names are underrepresented in the base data their machine learning process relies on. Therefore, for our analysis on gender, we use three categories: female, male, and other such that the last category contains all authorship records which do not belong to any of the first two categories.

Table 4.1: Five examples of affiliation series and the observed migration events in our model

Affiliation series	Modal countries of affiliation			Outcome in our model
	Year t_1	Year t_2	Year t_3	
E1	C_1	C_1	C_2	No migration is observed with the data for t_1 and t_2 because the mode country of affiliation has not changed between those years. A migration is observed later when the mode country of affiliation changes between t_2 and t_3.
E2	C_1		C_2	Migration from C_1 to C_2 is estimated to take place at $\lceil \frac{t_1+t_3}{2} \rceil$; taking the average is particularly useful for series of affiliations without any data points (publications) in some years.
E3	C_1	C_1 C_2	C_1 , C_2	No migration because the country of academic residence, C_1, is still among the mode countries in t_2 and t_3.
E4	C_1	C_1 ,C_2	C_2	Migration from C_1 to C_2 is observed between t_2 and t_3, therefore the migration year is recorded as $\lceil \frac{t_2+t_3}{2} \rceil$.
E5	C_1	C_2 ,C_3		A migration is observed between t_1 and t_2. The new country of academic residence is determined by a random selection between C_2 and C_3.

4.4 Descriptive Analysis

In order to understand the changing characteristics of brain circulation in the UK, first we employ a descriptive analysis. We visually explore the dynamic flows of researchers both moving to and from the UK, by academic origin, before and after Brexit; Figure 4.5 and 4.6 in the respectively. Then, we illustrate the trends of outgoing and incoming researchers by academic origin in the UK.

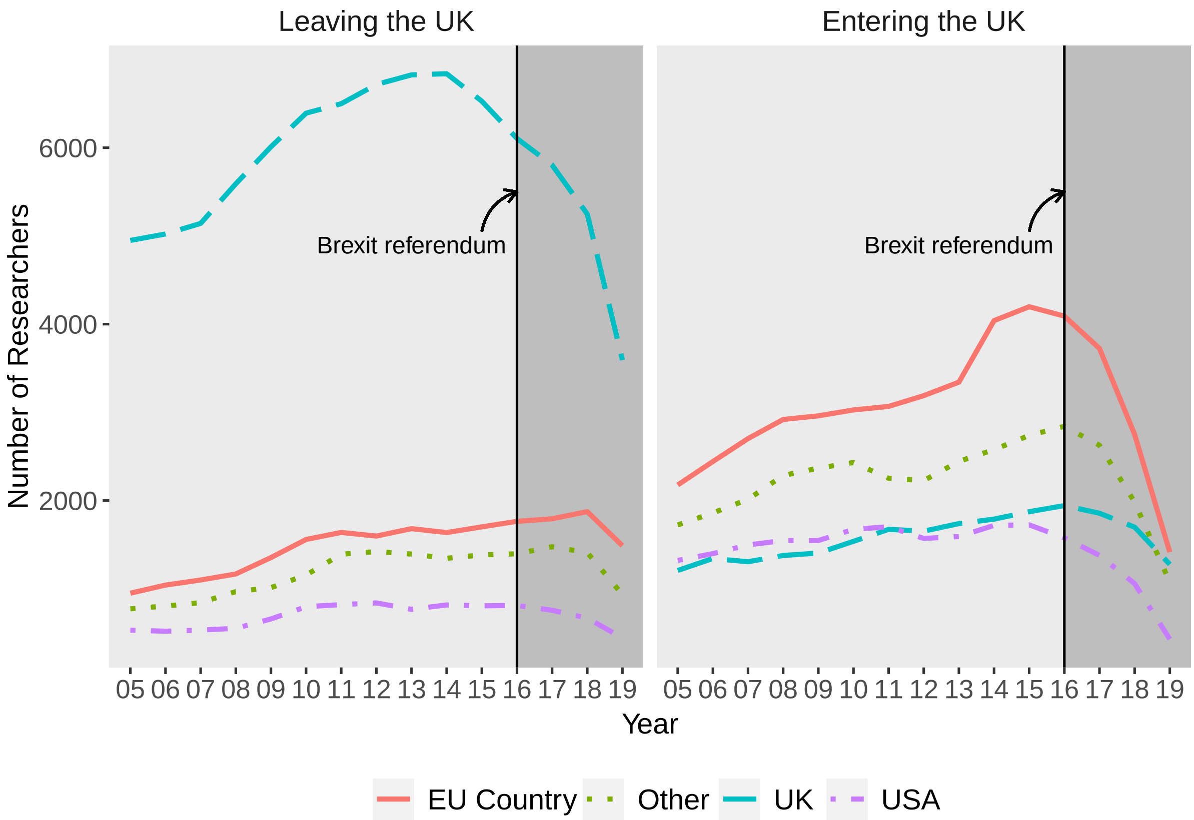

Figure 4.1: Researchers leaving and entering the UK, breakdown by academic origin, 2005-2019

In Figure 4.1, we see a declining trend for incoming researchers with the exception of a no-change observed for researchers with a Chinese academic origin. This decline after Brexit referendum would have important implications for British scientific environment, however, we have to be careful about jumping to conclusions regarding a decreasing trend, due to right-censoring in our data. Since not every researcher in our dataset publishes every year, such declines are observed for any random cut-offs, instead of the last year of available data. Therefore, the interesting trend in Figure 4.1 is not the decline in either incoming or outgoing migration depictions, but the increase of the trend for the outgoing researchers with an EU academic origin. That we observe an increase in the number of researchers whose academic origin is an EU country[1], despite the right-censoring and the underestimation of movements in the years close to the end year in our data (2019), points to the possibility of a Brexit-effect for a certain group of researchers residing in the UK.

Although we use the best possible data available, the nature of our dataset poses a challenge for empirical analysis due to the lack of observation in the years the authors have not published. To both avoid over-assuming no mobility in non-publishing years and to tackle this challenge, we focus on highly productive researchers. Thus, we introduce a third selection criteria, having at least one publication every year between 2013-2019, i.e. years surrounding the Brexit referendum. We define four academic origin categories as EU countries, USA, UK and other. The migration trends of highly productive researchers and the percentage shares of each academic origin category in overall outgoing and incoming researchers in UK are shown in Figure 4.2 and 4.3 respectively.

[1] The group of EU countries do not include the UK in this classification, neither before nor after Brexit.

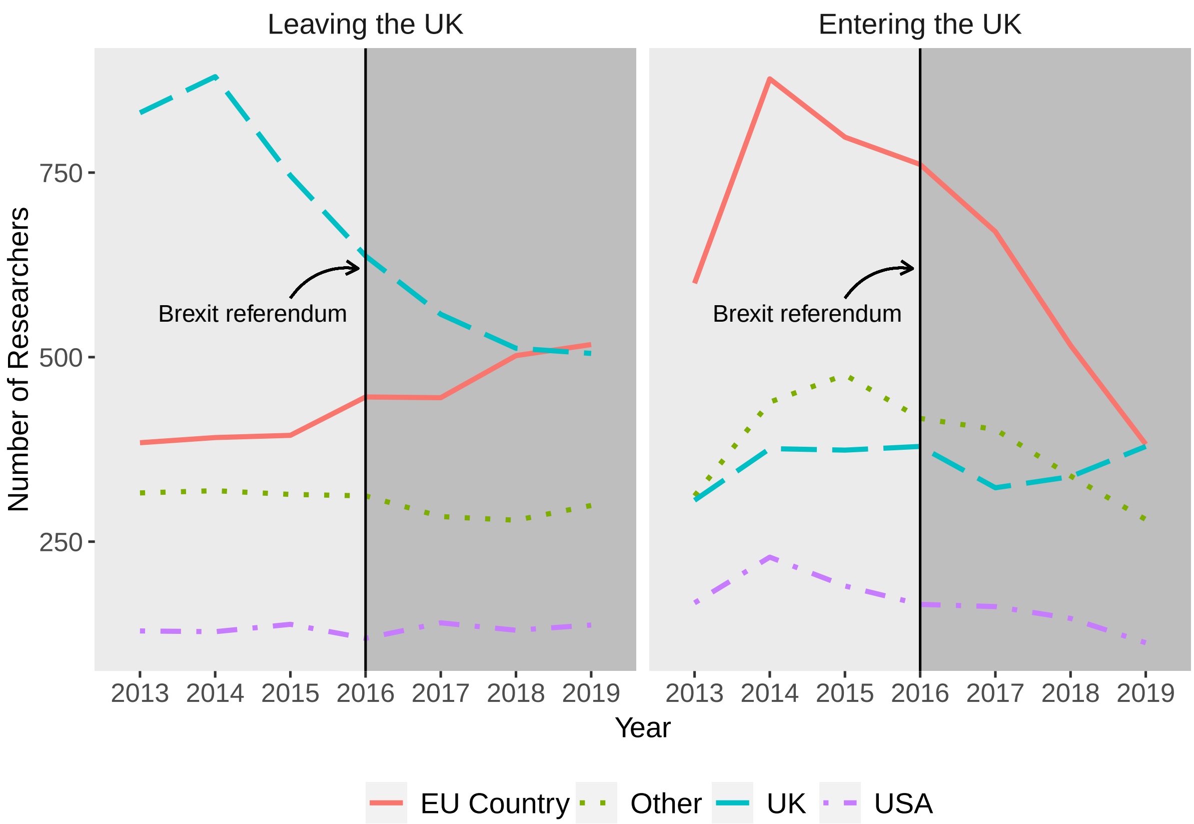

Figure 4.2: Highly productive researchers leaving and entering the UK, by academic origin (2013-2019)

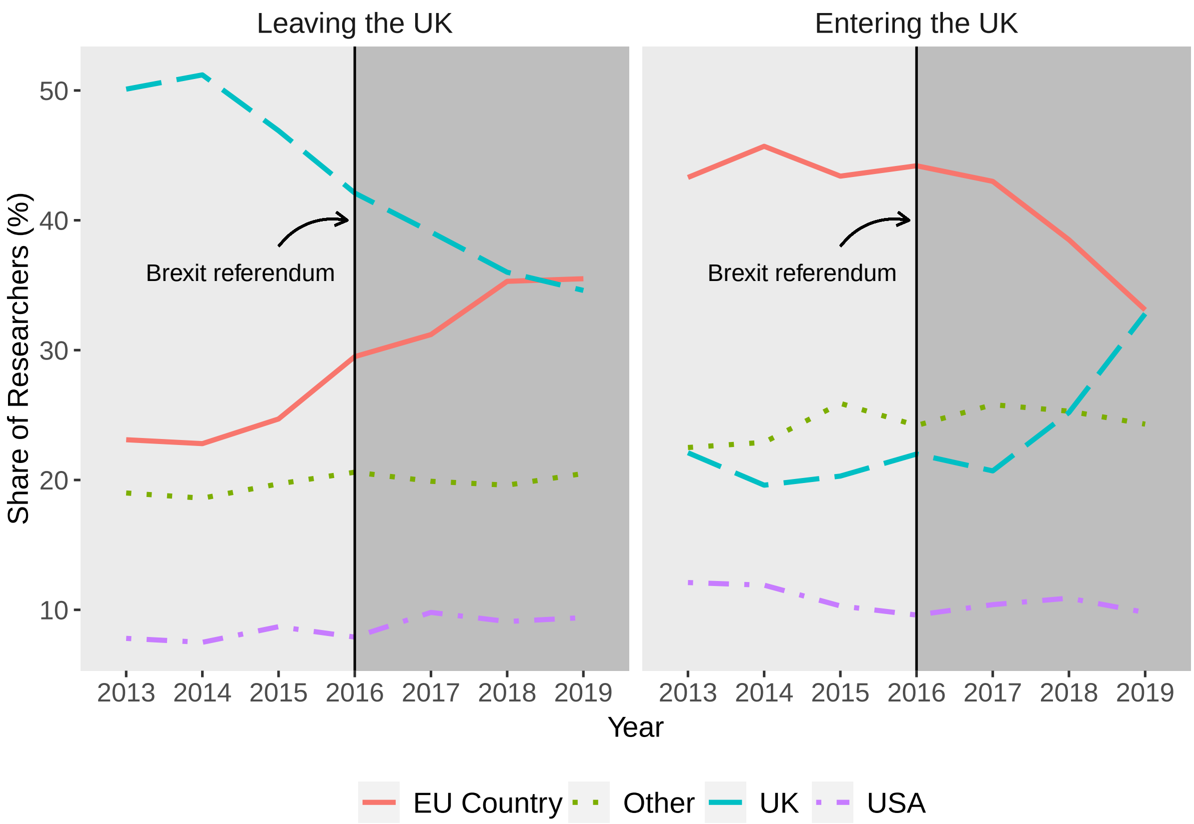

Figure 4.3: Share of academic origin in total highly productive researchers leaving and entering the UK (2013-2019)

Figures 4.2 and 4.3 depict the in- and out-migration patterns of highly productive researchers, i.e. researchers who have at least one publication every year between 2013-2019. As right-censoring is not an issue because we can observe the location and the change thereof for each researcher, for each year, interpretation of increases and decreases in migration trends are more accurate. We observe that the number of outgoing researchers whose academic origin is the UK (shown with the blue line) is in a decreasing trend, while there is an increase in the number of researchers with a UK academic origin, who return to the UK. In contrast, researchers with an EU academic origin are observed to increasingly leave the UK and are moving to the UK in less numbers. Same trends are observed, when their overall shares among incoming and outgoing researchers are taken into account.

4.5 Empirical Analysis

Considering the subset of data on highly productive researchers as more consistent, we organize this subset as panel data to use in the empirical analysis. Thus, to quantify the change in brain circulation patterns in the UK after Brexit, we employ a random effects logistic regression model using the panel data of highly productive researchers, consisting of 40.315 researchers and the years 2013-2019. The selection of the random effects model over fixed effects model relies on the intention to observe the effects of time-invariant characteristics, with academic origin being one of the main explanatory variables. We focus on the association of academic origin and years after Brexit referendum with the likelihood of leaving the UK, based on the model specifications below.

$$MovesOut_{i,t} = ln(P/1-P)_{i,t} = \alpha + \beta_1 Brexit * Origin_{i,t} + \beta_2 X_{i,t} + w_i + \tau_t + \epsilon_{i,t} \quad (4.1)$$

$$MovesIn_{i,t} = ln(P/1-P)_{i,t} = \alpha + \beta_1 Brexit * Origin_{i,t} + \beta_2 X_{i,t} + w_i + \tau_t + \epsilon_{i,t} \quad (4.2)$$

In the random effects logistic regression equation above, dependent variables $MovesOut$ and $MovesIn$ represent the binary variables equal to 1, when in a given year t, the researcher i leaves the UK and moves to the UK respectively. The main explanatory variables are denoted by the interaction term $Brexit*Origin$, while control variables are indicated by X. The variable $Brexit$ is a binary variable equal to 1 for years 2016-2019. Control variables include academic age, calculated by taking the year of first publication as academic birth, and dummy variables for having higher than average publication and citation count, scientific field (according to ASJC classification) and inferred gender.

The results of the random effects model stated in equations 4.1 and 4.2 are presented in Table 4.2 below. The plot of odds ratios depicting the association and significance of our variables of interest are further presented in Figure 4.4.

Table 4.2: Results of the random effects model

	Moving out of the UK		Moving into the UK	
VARIABLES	logit coefficient	odds ratio	logit coefficient	odds ratio
Brexit	0.324***	1.382***	0.268***	1.307***
	(0.0450)	(0.0621)	(0.0411)	(0.0537)
Origin: UK	0.270***	1.310***	-0.907***	0.404***
	(0.0375)	(0.0491)	(0.0409)	(0.0165)
Origin: EU	-0.236***	0.790***	0.189***	1.208***
	(0.0436)	(0.0344)	(0.0332)	(0.0401)
Origin: US	0.0147	1.015	0.181***	1.199***
	(0.0598)	(0.0606)	(0.0475)	(0.0569)
Brexit # Origin: UK	-0.349***	0.706***	0.168***	1.183***
	(0.0533)	(0.0376)	(0.0574)	(0.0679)
Brexit # Origin: EU	0.298***	1.347***	-0.0811	0.922
	(0.0584)	(0.0787)	(0.0511)	(0.0471)
Brexit # Origin: US	0.0574	1.059	-0.117	0.889
	(0.0811)	(0.0859)	(0.0725)	(0.0645)
Constant	-1.903***	0.149***	-1.662***	0.190***
	(0.108)	(0.0162)	(0.100)	(0.0190)
Controls	YES	YES	YES	YES
Observations	282,205	282,205	282,205	282,205
Number of researchers	40,315	40,315	40,315	40,315

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

In the random effects logit model highly productive researchers, whose academic origin is a country other than the US, UK or EU country, i.e. grouped as *other country*, are considered the baseline. The outcomes for the researchers with US, UK or EU country academic origins are reported relative to the baseline group. The results shown in Table 4.2 indicate a change in the mobility patterns for researchers whose academic origin is the UK or and EU country, following the Brexit referendum in 2016. Having an EU country as academic origin alone has a negative and statistically significant correlation with the chances of moving out of the UK, while the opposite is observed for the UK academic origin. However, when academic origin variable is interacted with the dummy variable of Brexit, taking the value of 1 after 2016, the chances of moving out is observed as reversed. In the post-Brexit context, having an EU country as an academic origin significantly increases the chances of moving out and the direction of the relationship turns from negative to positive with 55% increase in the odds of moving out. The association between chances of leaving the UK and having UK as the academic origin is also reversed in the post-Brexit context in comparison to the overall patterns and highly productive researchers with UK origin are observed as significantly less likely o leave the UK after Brexit. Furthermore, in terms of the chances of moving to the UK, we observe a contrast between general patterns and post-Brexit context for the UK academic origin. While highly productive researchers with UK academic origin are significantly less likely to return to UK in general, we observe that they become significantly more likely to move to the UK after Brexit.

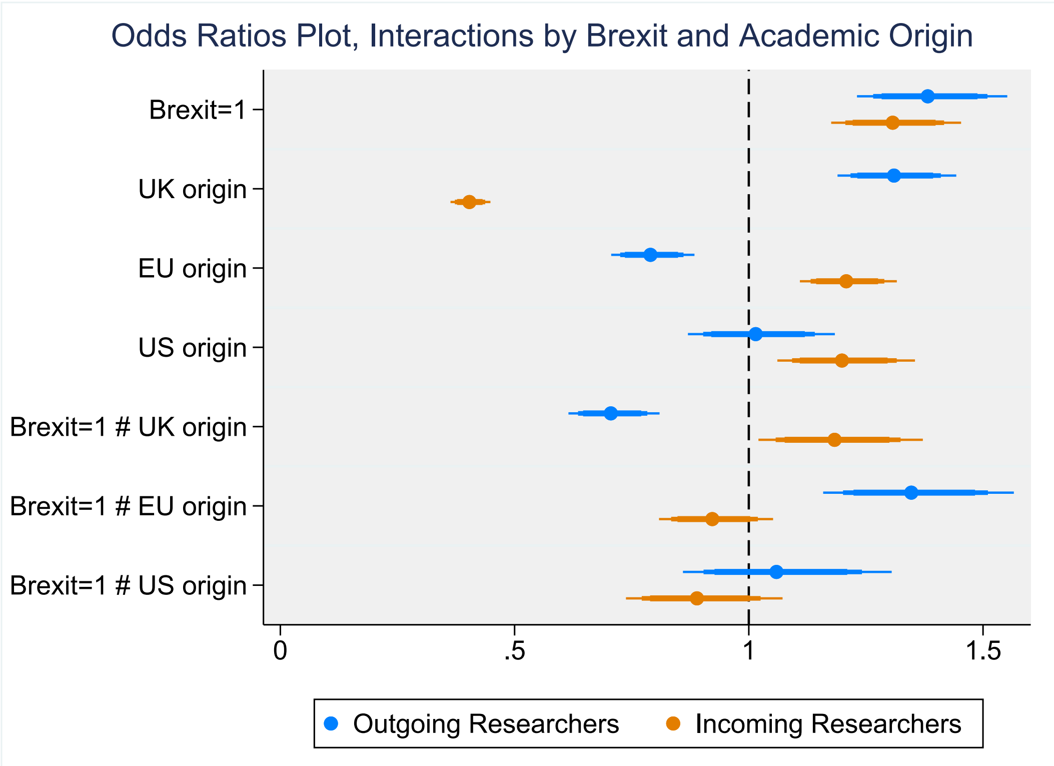

Figure 4.4: The odds ratios of leaving the UK for the highly productive researchers, by academic origin, between 2013-2019.

The odds ratio plot in Figure 4.4 illustrates the results reported in Table 4.2. In the figure, highly productive researchers moving out of the UK are shown in blue and highly productive researchers moving to the UK are shown in orange colours. The reversal of the migration patterns after Brexit referendum for the out-migration of highly productive researchers with UK and EU academic origins as well as the in-migration of highly productive researchers with UK academic origin are clearer in the plot. The figure also shows that the change in post-Brexit context is insignificant for highly productive researchers with US academic origin.

4.6 Conclusion

This study seeks to uncover the early influence of Brexit on British academia, using bibliometric data. The descriptive analysis of the bibliometric data suggests that, if the post-Brexit trends continue as it is, Brexit may trigger a change in the composition of the British scientific environment. The international mobility trends for highly productive researchers, defined as researchers with at least one publication per year between 2013-2019, demonstrate a stark contrast between pre- and post-Brexit periods, for researchers with UK or EU academic origin. In Figure 4.2 the number of highly productive outgoing researchers with EU academic origin show an increase following the Brexit referendum, while the number of highly productive incoming researchers with EU academic origin show a decline. The Figure 4.3 depicts the same picture for the shares of outgoing and incoming highly productive researchers by academic origin. Here, the change after Brexit referendum is the most clear for the incoming highly productive researchers, as the share of those with EU academic origin decrease while those with UK academic origin increase.

The empirical analysis we employ in the second part of the study corroborates with the descriptive analysis and confirms the statistical significance of the change also observed in descriptive analysis. The results of the random effects logistic regressions to assess the out-migration and in-migration patterns of highly productive researchers between 2013-2019 are presented in Table 4.2. As seen in Table 4.2, the odds of moving to the UK after Brexit is 18% higher for highly productive researchers with a UK academic origin in comparison to the baseline group of highly productive researchers with other academic origin. Without the condition for the move to occur after Brexit, the odds of moving to the UK is 60% lower for the highly productive researchers with a UK academic origin than the odds of moving to the UK for the baseline group.

Last, we demonstrate the odds ratios in Figure 4.4, where the interaction of post-Brexit move year and academic origin can be observed to reverse the international mobility patterns based on academic origin. Highly productive researchers with UK academic origin are significantly more likely to move to the UK and stay in the UK after Brexit, in stark contrast to their pre-Brexit migration patterns. Similarly, the plot shows that while highly productive researchers with EU academic origin are significantly less unlikely to leave the UK in general, when the interaction with post-Brexit years are taken into account, highly productive researchers with EU academic origin become significantly more likely to leave the UK.

The descriptive and empirical analyses in this chapter suggest that the scientific migration patterns in the UK is likely to be affected by the decision to withdraw from the European Union. However, the analyses do not indicate an overall escape of researchers from the UK after the Brexit, at least in the initial years after the referendum. Instead, our results imply that the change in the British scientific environment that followed the Brexit is more of a compositional change by academic origin. The results our analyses imply that highly productive researchers whose academic origin is an EU country are significantly more likely to leave the UK in stark contrast to their overall out-migration behaviour. In turn, highly productive researchers with UK academic origin become significantly less likely to leave the UK and more likely to return to the UK in the first years after the Brexit referendum, reversing their general migration patterns. The results indicate overall that the EU scientific institutions may benefit from the highly productive researchers leaving the UK. While in the initial years after Brexit, surrounded by uncertainty about the future of scientific funding, we do not observe a total outflow of highly productive researchers from the UK, the recent propensity of highly productive researchers with UK to leave the UK less and return to UK if abroad, bears the risk of reducing the scientific and institutional diversity in British scientific environment. Further research is needed to revisit these results a few years later to understand whether the changed migration behaviour by highly productive researchers with EU and UK academic origin will be long-lasting and, if so, to develop appropriate policy responses.

4.7 Appendix-Chapter 4

Table 4.3: Details of bibliometric data and what they indicate for scholarly migration

Scopus database	Indicator
Author-ID	Individual researcher
Country affiliation per publication, per Author-ID	Academic location
Country affiliation of the 1st publication per Author-ID	Country of academic origin
Publication year per publication per Author-ID	Year denoted by t_n
Publication year of the 1st publication per Author-ID	Academic birth year
Number of publications	Quantity-based measure of productivity
Mode of fields of publication venues for all publications per Author-ID	Main discipline of a researcher

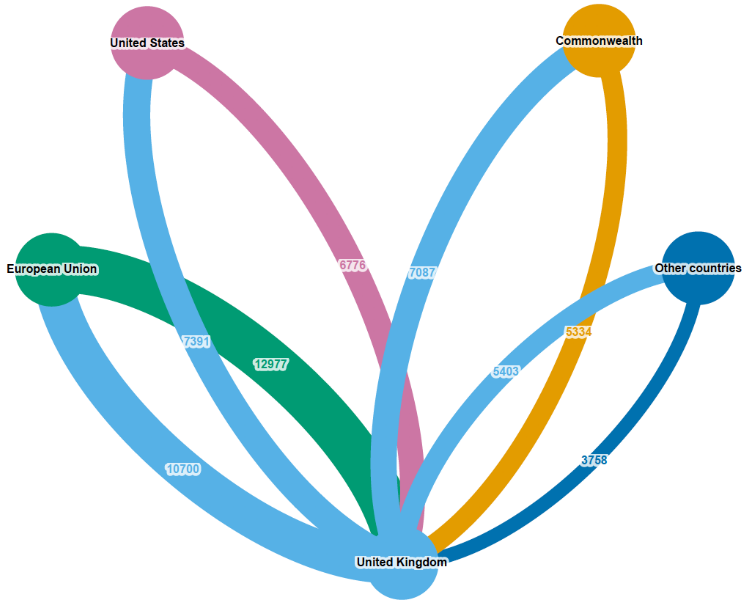

Figure 4.5: Migration flows and the patterns of scholarly migration, between 2013-2015. Migration flows and the overall patterns of scholarly migration in the three years prior to the Brexit referendum: EU having the largest flows to and from the UK followed by US, Commonwealth, and all *other* countries in a decreasing order. Edges represent the migration flows in 2013-2015. Direction of the edges are clock-wise. Colors of the edges are based on the origin node.

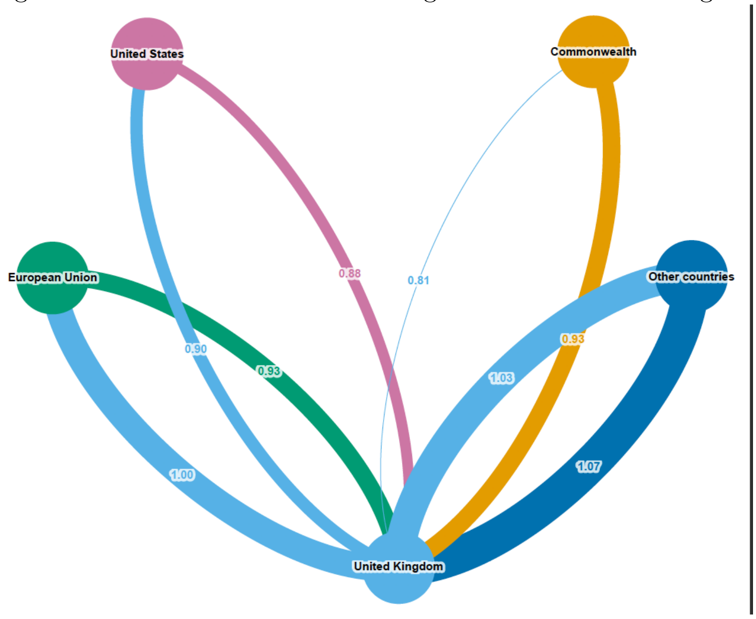

Figure 4.6: Migration flows and the patterns of scholarly migration, between 2016-2018 In the three years after the Brexit referendum, the flows have mostly decreased, except for the flow from the UK to EU (which has remained the same) and the flows between the UK and *other* countries (which have increase in both directions). Edges represent the changes in migration flows of 2016-2018 respective to 2013-2015. Direction of the edges are clock-wise. Colors of the edges are based on the origin node.

Chapter 5

Conclusion

This dissertation aims to contribute to the use of computational methods and digital data in social sciences, in specific demography and migration studies. The essays in this dissertation provides case studies for two different types of migration, forced migration & refugees in the Chapters 2 and 3, and the high-skilled migration in Chapter 4. Chapters 2 and 3 use Turkey and the Syrian citizens under temporary protection as the case study, while Chapter 4 focuses on an entirely different context, the British scientific environment, before and after Brexit. Chapters 2 and 3 employ quantitative analyses to understand the patterns of migration flows and concludes with a discussion of potential policy options. Chapter 4, in turn, departs from the introduction of a specific policy (the decision of the United Kingdom to leave the European Union) that has the potential to directly or indirectly influence high-skilled migration, and analyzes the early implications. In all these three chapters, digital data obtained by computational methods are used to different extents, along with the visualisation methods. The use of digital data, together with the volume of data being used intensify with each essay.

Chapter 2 provides a visual interpretation of the internal migration patterns in Turkey as well as the settlement patterns of Syrians under temporary protection in Turkey. The illustration of internal migration patterns in Turkey with a circular understanding, both contributes to the literature on circular visualization of migration and revisits the overall concept of circularity in migration patterns; i.e. enables to consider the migration currents with their counter-currents (Lee, 1966; Saxenian, 2005). Furthermore, the chapter questions whether replacement migration can provide a solution for sustainable refugee migration policy in Turkey. Using an empirical analysis, Chapter 2 finds that there is a negative and significant association between the migration patterns of locals and Syrians. It further shows that the Syrians in Turkey choose to settle in provinces where they have either cultural & linguistic ties or job opportunities. Last, but not least, the chapter offers a contribution to the social science studies on Turkey in terms of the use of native language data. By showing that there is a significant change in the first and last available data on native tongues in Turkey, the essay suggests that use of 1927 data, before the relative establishment of union in language, may be more useful than the use of 1965 data.

Chapter 3 of the dissertation also focuses on the case of Syrians under temporary protection in Turkey however, it seeks to understand the mobility of Syrians under temporary protection inside Turkey, using Google Trends data as proxy. In data collection, the second chapter also makes use of innovative approaches as *Wayback Machine* and *WebCite*. The results of Chapter 3 demonstrates that there is a positive and significant association between the Google searches made for province names using Arabic letters and the refugee stock of provinces. The association is significant, if the search has been made one-month and six-weeks prior to the observation of respective change in official statistics. This result is also coherent with the time needed for preparation before travel and necessary bureaucratic steps. The findings of this chapter contributes to the literature on forecasting and in specific the literature using Google Trends data as an indicator. Furthermore, the findings show that forecasting the movements of Syrians uTP to different provinces, and migrants in general, is possible using this method, which may help local and national governments in developing policies for migration management.

Chapter 4 of this dissertation focuses on the international mobility of scientists as a case of high-skilled migration. Using the UK scientific environment as the context and Brexit as the potential policy effect, the chapter seeks to reveal the initial influence of Brexit on researchers' migration using a big data approach with bibliometric data. The findings of the chapter show that among the UK-resident scholars, the odds of leaving the UK after Brexit is 35% higher if the academic origin is an EU country, while the odds are 21% lower for the same group of scholars without the Brexit condition. Furthermore, the odds of returning to the UK is 18% higher for scholars with UK academic origin after Brexit, in contrast to their pre-Brexit migration patterns. The findings of the Chapter 4 also contributes to the literature on brain circulation (Saxenian, 2005) and corroborates with the literature that high-skilled migrants are affected by potential incentives for migration in different ways. In this case, the policy of Brexit does not seem to initiate a brain drain out of the UK but a compositional change in British academic environment. While this change may create issues due to lack of diversity, it is a different outcome than the brain drain and requires different policies to be addressed.

Bibliography

Guy J Abel and Nayoung Heo. Changing internal migration flows patterns in south korea. *Regional Studies, Regional Science*, 5(1):78–80, 2018.

Guy J Abel and Nikola Sander. Quantifying global international migration flows. *Science*, 343 (6178):1520–1522, 2014.

Louise Ackers. Moving People and Knowledge: Scientific Mobility in the European Union1. *International Migration*, 43(5):99–131, 2005. ISSN 1468-2435. doi: 10.1111/j.1468-2435.2005. 00343.x.

Jonathan Adams, Gordon Rogers, and Martin Szomszor. *The Annual G20 Scorecard: Research Performance 2019*. Institute for Scientific Information, 2019.

AFAD. Türkiye'deki suriyelilerin demografik görünümü, yaşam koşulları ve gelecek beklentilerine yönelik saha araştırması [field research on the demographic outlook of syrians in turkey, life conditions and expectations from the future], 2017.

Ajay K Agrawal, John McHale, and Alex Oettl. Why stars matter. Technical report, National Bureau of Economic Research, 2014.

Diego Alburez-Gutierrez, Samin Aref, Sofia Gil-Clavel, Andre Grow, Daniela Negraia, and Emilio Zagheni. Demography in the digital era: New data sources for population research. In G. Arbia, S. Peluso, A. Pini, and G. Rivellini, editors, *Book of Short Papers SIS2019*, pages 23–30. Pearson, 2019. doi: 10.31235/osf.io/24jp7.

Onur Altındağ and Neeraj Kaushal. Do refugees impact voting behavior in the host country? evidence from syrian refugee inflows to turkey. *Public Choice*, pages 1–30, 2020.

Anonymous. Gaps and drains. *Nature*, 214:223, 1967/04. doi: \url{https://doi.org/10.1038/214223a0}.

Samin Aref, David Friggens, and Shaun Hendy. Analysing scientific collaborations of New Zealand institutions using Scopus bibliometric data. In *Proceedings of the Australasian Computer Science Week Multiconference*, page 49, New York, NY, USA, 2018. ACM. doi: 10.1145/3167918.3167920.

Samin Aref, Emilio Zagheni, and Jevin West. The demography of the peripatetic researcher: Evidence on highly mobile scholars from the web of science. In Ingmar Weber, Kareem M.

Darwish, Claudia Wagner, Emilio Zagheni, Laura Nelson, Samin Aref, and Fabian Flöck, editors, *Social Informatics*, pages 50–65. Springer International Publishing, 2019. ISBN 978-3-030-34971-4.

Sanjay K Arora, Yin Li, Jan Youtie, and Philip Shapira. Using the wayback machine to mine websites in the social sciences: a methodological resource. *Journal of the Association for Information Science and Technology*, 67(8):1904–1915, 2016.

Omar A Doria Arrieta, Fabio Pammolli, and Alexander M Petersen. Quantifying the negative impact of brain drain on the integration of european science. *Science advances*, 3(4):e1602232, 2017.

Nikos Askitas and Klaus F Zimmermann. Google econometrics and unemployment forecasting. *IZA Discussion Papers, No. 4201*, 2009.

G Atalik and H Ciraci. Bolgelerarasi farklilasmanin bir sonucu olarak bolgelerarasi gocler. [interregional migration as a result of interregional divergence]. *3. Ulusal Bölge Bilimi/Bölge Planlama Kongresi, Istanbul*, pages 109–119, 1993.

Pierre Azoulay, Ina Ganguli, and Joshua Graff Zivin. The mobility of elite life scientists: Professional and personal determinants. *Research Policy*, 46(3):573 – 590, 2017. ISSN 0048-7333. doi: https://doi.org/10.1016/j.respol.2017.01.002.

Kirk Bansak, Jeremy Ferwerda, Jens Hainmueller, Andrea Dillon, Dominik Hangartner, Duncan Lawrence, and Jeremy Weinstein. Improving refugee integration through data-driven algorithmic assignment. *Science*, 359(6373):325–329, 2018.

Emerson Augusto Baptista, Guy J Abel, and Járvis Campos. Internal migration in brazil using circular visualization. *Regional Studies, Regional Science*, 5(1):361–364, 2018.

Michel Beine, Frédéric Docquier, and Hillel Rapoport. Brain drain and economic growth: theory and evidence. *Journal of development economics*, 64(1):275–289, 2001.

Maureen Benson-Rea and Stephen Rawlinson. Highly skilled and business migrants: Information processes and settlement outcomes. *International Migration*, 41(2):59–79, 2003.

Jagdish Bhagwati and Koichi Hamada. The brain drain, international integration of markets for professionals and unemployment: a theoretical analysis. *Journal of Development Economics*, 1(1):19–42, 1974.

Francesco Billari, Francesco D'Amuri, and Juri Marcucci. Forecasting births using google. In *CARMA 2016-1st International Conference on Advanced Research Methods and Analytics*, 2016.

Francesco C Billari, Valentina Rotondi, and Jenny Trinitapoli. Mobile phones, digital inequality, and fertility. *Demographic Research*, 42:1057–1096, 2020.

Albert Bollard, David McKenzie, Melanie Morten, and Hillel Rapoport. Remittances and the brain drain revisited: The microdata show that more educated migrants remit more. *The World Bank Economic Review*, 25(1):132–156, 2011.

Abel Brodeur, Andrew Clark, Sarah Fleche, and Nattavudh Powdthavee. Covid-19, lockdowns and well-being: Evidence from google trends. 2020.

John S Brownstein, Clark C Freifeld, and Lawrence C Madoff. Digital disease detection—harnessing the web for public health surveillance. *The New England journal of medicine*, 360(21):2153, 2009.

Elizabeth E Bruch and MEJ Newman. Aspirational pursuit of mates in online dating markets. *Science Advances*, 4(8):eaap9815, 2018.

Nina Cesare, Hedwig Lee, Tyler McCormick, Emma Spiro, and Emilio Zagheni. Promises and pitfalls of using digital traces for demographic research. *Demography*, 55(5):1979–1999, 2018.

Elizabeth Chacko. From brain drain to brain gain: reverse migration to bangalore and hyderabad, india's globalizing high tech cities. *GeoJournal*, 68(2):131–140, 2007.

Shu-Sen Chang, Simon Sai Man Kwok, Qijin Cheng, Paul SF Yip, and Ying-Yeh Chen. The association of trends in charcoal-burning suicide with google search and newspaper reporting in taiwan: a time series analysis. *Social psychiatry and psychiatric epidemiology*, 50(9):1451–1461, 2015.

Elin Charles-Edwards, Tom Wilson, and Nikola Sander. Visualizing australian internal and international migration flows. *Regional Studies, Regional Science*, 2(1):432–434, 2015.

Barry R Chiswick. High skilled immigration in the international arena. 2005.

Hyunyoung Choi and Hal Varian. Predicting initial claims for unemployment benefits. *Google Inc*, 1:1–5, 2009.

Michael A Clemens. Does development reduce migration? In *International Handbook on migration and Economic development*. Edward Elgar Publishing, 2014.

Phillip Connor. The digital footprint of europe's refugees, 2017.

Pedro Cruz, John Wihbey, Avni Ghael, Felipe Shibuya, and Stephen Costa. Dendrochronology of us immigration. *Information Design Journal*, 25(1):6–20, 2019.

Mathias Czaika. *High-skilled Migration: Drivers and Policies*. Oxford University Press, New York, NY, USA, 2018. ISBN 978-0-19-185302-9. doi: 10.1093/oso/9780198815273.001.0001.

Mathias Czaika and Christopher R Parsons. The gravity of high-skilled migration policies. *Demography*, 54(2):603–630, 2017.

Hein De Haas. International migration, remittances and development: myths and facts. *Third World Quarterly*, 26(8):1269–1284, 2005.

Hein De Haas. Turning the tide? why development will not stop migration. *Development and change*, 38(5):819–841, 2007.

Hein De Haas. Migration transitions, 2010.

Rianne Dekker and Godfried Engbersen. How social media transform migrant networks and facilitate migration. *Global Networks*, 14(4):401–418, 2014.

Rianne Dekker, Godfried Engbersen, Jeanine Klaver, and Hanna Vonk. Smart refugees: How syrian asylum migrants use social media information in migration decision-making. *Social Media+ Society*, 4(1):2056305118764439, 2018.

Dana Diminescu. The connected migrant: an epistemological manifesto. *Social Science Information*, 47(4):565–579, 2008.

Frédéric Docquier and Hillel Rapoport. Globalization, brain drain, and development. *Journal of economic literature*, 50(3):681–730, 2012.

Frédéric Docquier, Hillel Rapoport, and Sara Salomone. Remittances, migrants' education and immigration policy: Theory and evidence from bilateral data. *Regional Science and Urban Economics*, 42(5):817–828, 2012.

Mehmet Ali Eryurt and İsmet Koç. Demography of ethnicity in turkey. In *The International Handbook of the Demography of Race and Ethnicity*, pages 483–502. Springer, 2015.

Michael Ettredge, John Gerdes, and Gilbert Karuga. Using web-based search data to predict macroeconomic statistics. *Communications of the ACM*, 48(11):87–92, 2005.

Beatrix Eugster, Rafael Lalive, Andreas Steinhauer, and Josef Zweimüller. Culture, work attitudes, and job search: Evidence from the swiss language border. *Journal of the European Economic Association*, 15(5):1056–1100, 2017.

European Commission. H2020 country profile: Key figures - United Kingdom. `https://webgate.ec.europa.eu/dashboard/sense/app/a976d168-2023-41d8-acec-e77640154726/sheet/0c8af38b-b73c-4da2-ba41-73ea34ab7ac4/state/0`, 2020.

Alessandra Faggian and Rachel S Franklin. Human capital redistribution in the usa: the migration of the college-bound. *Spatial Economic Analysis*, 9(4):376–395, 2014.

Ana Fernández-Zubieta, Aldo Geuna, and Cornelia Lawson. Productivity pay-offs from academic mobility: should i stay or should i go? *Industrial and Corporate Change*, 25(1):91–114, 2016.

Alpay Filiztekin and Ali Gökhan. The determinants of internal migration in turkey. In *International Conference on Policy Modelling (EcoMod2008), Berlin, Germany*, volume 24, 2008.

Chiara Franzoni, Giuseppe Scellato, and Paula Stephan. International mobility of research scientists: Lessons from GlobSci. In *Global Mobility of Research Scientists*, pages 35–65. Elsevier, 2015. doi: 10.1016/B978-0-12-801396-0.00002-8.

Carolos Frenk, Tim Hunt, L Partridge, Janet Thornton, and Terry Wyatt. Uk research and the european union: The role of the eu in funding uk research. *The Royal Society*, pages 1–31, 2015.

Ferhan Gezici and Geoffrey JD Hewings. Regional convergence and the economic performance of peripheral areas in turkey. *Review of Urban & Regional Development Studies*, 16(2):113–132, 2004.

Ferhan Gezici and Berna Keskin. Interaction between regional inequalities and internal migration in turkey. In *ERSA Conference Papers*, pages 1–18, 2005.

Sofia Gil-Clavel and Emilio Zagheni. Demographic differentials in facebook usage around the world. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 13, pages 647–650, 2019.

Marie Gillespie, Souad Osseiran, and Margie Cheesman. Syrian refugees and the digital passage to europe: Smartphone infrastructures and affordances. *Social Media+ Society*, 4(1): 2056305118764440, 2018.

Jeremy Ginsberg, Matthew H Mohebbi, Rajan S Patel, Lynnette Brammer, Mark S Smolinski, and Larry Brilliant. Detecting influenza epidemics using search engine query data. *Nature*, 457(7232):1012–1014, 2009.

Matthew Godwin, Jane Gregory, and Brian Balmer. The Anatomy of the Brain Drain Debate, 1950–1970s: Witness Seminar. *Contemporary British History*, 23(1):35–60, March 2009. ISSN 1361-9462, 1743-7997. doi: 10.1080/13619460801990088.

Bernard T Golding. The impact of Brexit on British science: Also Sprach Zunderland. *Future Medicinal Chemistry*, 9(3):271–273, 2017. ISSN 1756-8919. doi: 10.4155/fmc-2016-0245.

Google. Faq about google trends data. `https://support.google.com/trends/answer/4365533?hl=en&ref_topic=6248052`, 2020. Online, Accessed November 2020.

Zuguang Gu, Lei Gu, Roland Eils, Matthias Schlesner, and Benedikt Brors. circlize implements and enhances circular visualization in r. *Bioinformatics*, 30(19):2811–2812, 2014.

Douglas T Gurak and Fee Caces. Migration networks and the shaping of migration systems. *International migration systems: A global approach*, pages 150–176, 1992.

Giselle Guzman. Internet search behavior as an economic forecasting tool: The case of inflation expectations. *Journal of economic and social measurement*, 36(3):119–167, 2011.

Nadeem U Haque and Se-Jik Kim. "human capital flight": Impact of migration on income and growth. *Staff Papers*, 42(3):577–607, 1995.

B. Hawelka, I. Sitko, E. Beinat, S. Sobolevsky, P. Kazakopoulos, and C. Ratti. Geo-located twitter as proxy for global mobility patterns. *Cartography and Geographic Information Science*, 41(3):260–271, 2014.

Metka Hercog and Laure Sandoz. Highly skilled or highly wanted migrants? conceptualizations, policy designs and implementations of high-skilled migration policies, 2018.

Rosalind S Hunter, Andrew J Oswald, and Bruce G Charlton. The elite brain drain. *The Economic Journal*, 119(538):F231–F251, 2009.

Turkish Statistical Institute. Survey on information and communication technology (ict) usage survey in households and by individuals, 2018.

Turkish Statistical Institute. Address based population registration system - scope of the data, 2020. Online; accessed Nov. 2020.

IOM. Glossary on migration, 2019.

John Irvine, Ben R. Martin, Tim Peacock, and Roy Turner. Charting the decline in British science. *Nature*, 316(6029):587–590, 1985. ISSN 1476-4687. doi: 10.1038/316587a0.

Harry G. Johnson. The economics of the "brain drain": The Canadian case. *Minerva*, 3(3): 299–311, 1965. ISSN 0026-4695.

Kemal Kirişci and Elizabeth G Ferris. *Not likely to go home: Syrian refugees and the challenges to Turkey-and the international community*. Brookings Institution Press, 2015.

Ismet Koc, Attila Hancioglu, and Alanur Cavlin. Demographic differentials and demographic integration of turkish and kurdish populations in turkey. *Population research and policy review*, 27(4):447–457, 2008.

Vincent Larivière, Chaoqun Ni, Yves Gingras, Blaise Cronin, and Cassidy R. Sugimoto. Bibliometrics: Global gender disparities in science. *Nature News*, 504(7479):211, 2013. doi: 10.1038/504211a.

Grit Laudel. Studying the brain drain: Can bibliometric methods help? *Scientometrics*, 57(2): 215–237, 2003. doi: 10.1023/A:1024137718393.

Grit Laudel. Migration Currents Among the Scientific Elite. *Minerva*, 43(4):377–395, 2005. ISSN 1573-1871. doi: 10.1007/s11024-005-2474-7.

Everett S Lee. A theory of migration. *Demography*, 3(1):47–57, 1966.

Danny M Leipziger. "brain drain" and the global mobility of high-skilled talent. 2008.

Allen Yilun Lin, Justin Cranshaw, and Scott Counts. Forecasting us domestic migration using internet search queries. In *The World Wide Web Conference*, pages 1061–1072, 2019.

Sami Mahroum. Highly skilled globetrotters: mapping the international migration of human capital. *R&D Management*, 30(1):23–32, 2000.

Sami Mahroum. The international policies of brain gain: a review. *Technology Analysis & Strategic Management*, 17(2):219–230, 2005.

Sarah Main. Here's how we can be a science superpower outside of the eu. *The Telegraph*, 2018.

Ben R. Martin. British science in the 1980s — Has the relative decline continued? *Scientometrics*, 29(1):27–56, 1994. ISSN 0138-9130. doi: 10.1007/BF02018382.

Ben R. Martin and John Irvine. Assessing basic research: Some partial indicators of scientific progress in radio astronomy. *Research Policy*, 12(2):61–90, 1983. ISSN 0048-7333. doi: 10.1016/0048-7333(83)90005-7.

Ben R. Martin, John Irvine, Francis Narin, and Chris Sterritt. The continuing decline of British science. *Nature*, 330(6144):123–126, 1987. ISSN 1476-4687. doi: 10.1038/330123a0.

Douglas S Massey and Felipe García España. The social process of international migration. *Science*, 237(4816):733–738, 1987.

Michael J McCarthy. Internet monitoring of suicide risk in the population. *Journal of affective disorders*, 122(3):277–279, 2010.

Letizia Mencarini, Delia Irazú Hernández-Farías, Mirko Lai, Viviana Patti, Emilio Sulis, and Daniele Vignoli. Happy parents' tweets. *Demographic Research*, 40:693–724, 2019.

Jean-Baptiste Meyer. Network approach versus brain drain: lessons from the diaspora. *International migration*, 39(5):91–110, 2001.

Jean-Baptiste Meyer, Jorge Charum, Dora Bernal, Jacques Gaillard, José Granés, John Leon, Alvaro Montenegro, Alvaro Morales, Carlos Murcia, Nora Narvaez-Berthelemot, et al. Turning brain drain into brain gain: the colombian experience of the diaspora option. *Science, technology and Society*, 2(2):285–315, 1997.

John Mingers and Loet Leydesdorff. A review of theory and practice in scientometrics. *European journal of operational research*, 246(1):1–19, 2015.

Kaz Miyagiwa. Scale economies in education and the brain drain problem. *International Economic Review*, pages 743–759, 1991.

H. F. Moed, W. J. M. Burger, J. G. Frankfort, and A. F. J. Van Raan. The use of bibliometric data for the measurement of university research performance. *Research Policy*, 14(3):131–149, 1985. ISSN 0048-7333. doi: 10.1016/0048-7333(85)90012-5.

Henk F. Moed and Gali Halevi. A bibliometric approach to tracking international scientific migration. *Scientometrics*, 101(3):1987–2001, 2014. doi: 10.1007/s11192-014-1307-6.

Henk F. Moed, M'hamed Aisati, and Andrew Plume. Studying scientific migration in Scopus. *Scientometrics*, 94(3):929–942, 2013. ISSN 1588-2861. doi: 10.1007/s11192-012-0783-9.

Petra Moser, Alessandra Voena, and Fabian Waldinger. German jewish émigrés and us invention. *American Economic Review*, 104(10):3222–55, 2014.

Gabriela F. Nane, Vincent Larivière, and Rodrigo Costas. Predicting the age of researchers using bibliometric data. *Journal of Informetrics*, 11(3):713–729, 2017. ISSN 1751-1577. doi: 10.1016/j.joi.2017.05.002.

Max Nathan. The wider economic impacts of high-skilled migrants: a survey of the literature for receiving countries. *IZA Journal of Migration*, 3(1):1–20, 2014.

Denis Noble. From 'Save British Science', to 'CaSE', to the future. `http://www.sciencecampaign.org.uk/news-media/guest-blog/from-save-british-science-to-case-to-the-future.html`, 2016.

UN Population Division (UN Department of Economic and Social Affairs). Replacement migration: Is it a solution to ageing and declining populations, 2001.

Directorate General of Migration Management. Temporary protection statistics. URL: `https://www.goc.gov.tr/gecici-koruma5638`, 2020. Online; accessed Nov. 2020.

Richard C. Oldfield, J. A. Simmons, J. W. Jeffery, W. Mansfield Cooper, R. J. Eden, G. O. Jones, V. Kondic, John Mcmichael, E. R. Pike, K. W. Andrews, W. L. Bragg, D. M. S. Bagguley, J. M. Baker, A. H. Cooke, R. J. Elliott, J. H. E. Griffiths, D. ter Háar, J. Hatton, R. W. Hill, J. T. Houghton, H. G. Kuhn, N. Kurti, F. N. H. Robinson, B. V. Rollin, H. M. Rosenberg, J. H. Sanders, G. W. Series, J. Wilks, G. K. Woodgate, G. W. Hutchinson, S. K. Runcorn, T. GOLD, Norman Albon, Alexander Haddow, R. M. Hartwell, John Holmes, P. J. Garrett, S. Garrett, Peter F. Knowles, Gerard P. Moss, A. Prince, G. Sachs, David Weitzman, Alan R. Williamson, and Paul Holliday. Emigration of scientists from the United Kingdom: Report of a committee appointed by the council of the Royal Society. *Minerva*, 1(3):358–380, 1963.

Said Ouaked. Transatlantic roundtable on high–skilled migration and sending countries issues. *International Migration*, 40(4):153–166, 2002.

John RB Palmer, Thomas J Espenshade, Frederic Bartumeus, Chang Y Chung, Necati Ercan Ozgencil, and Kathleen Li. New approaches to human mobility: Using mobile phones for demographic research. *Demography*, 50(3):1105–1128, 2013.

Constantine Passaris. Immigration and the evolution of economic theory. *International migration (Geneva, Switzerland)*, 27(4):525–542, 1989.

Camille Pelat, Clement Turbelin, Avner Bar-Hen, Antoine Flahault, and Alain-Jacques Valleron. More diseases tracked by using google trends. *Emerging infectious diseases*, 15(8):1327, 2009.

Wei Qi, Guy J Abel, Raya Muttarak, and Shenghe Liu. Circular visualization of china's internal migration flows 2010–2015. *Environment and Planning A: Economy and Space*, 49(11):2432–2436, 2017.

Ernst Georg Ravenstein. The laws of migration. *Journal of the statistical society of London*, 48 (2):167–235, 1885.

Mark Regets. Research and policy issues in high-skilled international migration: A perspective with data from the united states. *Innovative people: Mobility of skilled personnel in national innovation systems*, pages 243–260, 2001.

Temporary Protection Regulation. Tc official gazette, 29153, 2014.

Ben Y Reis and John S Brownstein. Measuring the impact of health policies using internet search patterns: the case of abortion. *BMC public health*, 10(1):514, 2010.

Comfort F Ricketts and Carlos G Silva. An analysis of morbidity and mortality using google trends. *Journal of Human Behavior in the Social Environment*, 27(6):559–570, 2017.

Nicolás Robinson-García, Carolina Cañibano, Richard Woolley, and Rodrigo Costas. Scientific mobility of early career researchers in Spain and the Netherlands through their publications. In *21st International Conference on Science and Technology Indicators-STI 2016. Book of Proceedings*, Valencia, Spain, 2016. doi: 10.4995/STI2016.2016.4543.

Nicolás Robinson-García, Cassidy R. Sugimoto, Dakota Murray, Alfredo Yegros-Yegros, Vincent Larivière, and Rodrigo Costas. The many faces of mobility: Using bibliometric data to measure the movement of scientists. *Journal of Informetrics*, 13(1):50–63, 2019. doi: 10.1016/j.joi.2018.11.002.

Albert Ali Salah, Alex Pentland, Bruno Lepri, Emmanuel Letouzé, Patrick Vinck, Yves-Alexandre de Montjoye, Xiaowen Dong, and Ozge Dagdelen. Data for refugees: the d4r challenge on mobility of syrian refugees in turkey. *arXiv preprint arXiv:1807.00523*, 2018.

AnnaLee Saxenian. From brain drain to brain circulation: Transnational communities and regional upgrading in india and china. *Studies in comparative international development*, 40 (2):35–61, 2005.

Stefano Sbalchiero and Arjuna Tuzzi. Italian scientists abroad in europe's scientific research scenario: High skill migration as a resource for development in italy. *International Migration*, 55(4):171–187, 2017.

Giuseppe Scellato, Chiara Franzoni, and Paula Stephan. Migrant scientists and international networks. *Research Policy*, 44(1):108–120, 2015. ISSN 0048-7333. doi: 10.1016/j.respol.2014.07.014.

Paola Solano, Morena Ustulin, Enrico Pizzorno, Monica Vichi, Maurizio Pompili, Gianluca Serafini, and Mario Amore. A google-based approach for monitoring suicide risk. *Psychiatry research*, 246:581–586, 2016.

Tae Min Song, Juyoung Song, Ji-Young An, Laura L Hayman, and Jong-Min Woo. Psychological and social factors affecting internet searches on suicide in korea: a big data analysis of google search trends. *Yonsei medical journal*, 55(1):254–263, 2014.

Oded Stark, Christian Helmenstein, and Alexia Prskawetz. Human capital depletion, human capital formation, and migration: a blessing or a "curse"? *Economics Letters*, 60(3):363–367, 1998.

Bogdan State, Mario Rodriguez, Dirk Helbing, and Emilio Zagheni. Migration of professionals to the us, 2014.

Alexander Subbotin and Samin Aref. Brain drain and brain gain in Russia: Analyzing international mobility of researchers by discipline using Scopus bibliometric data 1996-2020. *ArXiv*, 2020.

Cassidy R Sugimoto and Vincent Larivière. *Measuring Research: What Everyone Needs to Know*. Oxford University Press, New York, NY, USA, 2018.

Cassidy R Sugimoto, Nicolás Robinson-García, and Rodrigo Costas. Towards a global scientific brain: Indicators of researcher mobility using co-affiliation data. *arXiv preprint arXiv:1609.06499*, 2016.

Cassidy R. Sugimoto, Nicolás Robinson-García, Dakota S. Murray, Alfredo Yegros-Yegros, Rodrigo Costas, and Vincent Larivière. Scientists have most impact when they're free to move. *Nature News*, 550(7674):29, 2017. doi: 10.1038/550029a.

Ulas Sunata. Syrian refugee livelihood monitor: summary assessment. *Syrian Refugee Livelihood Monitor*, pages 1–14, 2017.

J Edward Taylor. Differential migration, networks, information and risk. *Migration, human capital and development*, 4:147–171, 1986.

The GSM Association. The importance of mobile for refugees: A landscape of new services and approaches. URL: `https://www.gsma.com/mobilefordevelopment/wp-content/uploads/2017/02/The-Importance-of-mobile-for-refugees_a-landscape-of-new-services-and-approaches.pdf`, 2017. Online; accessed Nov. 2020.

Reuben J Thomas. Online exogamy reconsidered: Estimating the internet's effects on racial, educational, religious, political and age assortative mating. *Social Forces*, 98(3):1257–1286, 2020.

I Ulutürk, I Uysal, and O Varol. Refugee integration in turkey: a study of mobile phone data for d4r challenge. In *Data for refugees challenge workshop*, 2019.

UNHCR. Global trends: forced displacement in 2017, 2018.

Kamil Wais. Gender Prediction Methods Based on First Names with genderizeR. *The R Journal*, 8(1):17–37, 2006. doi: 10.32614/RJ-2016-002.

Susumu Watanabe. The brain drain from developing to developed countries. *Int'l Lab. Rev.*, 99:401, 1969.

Ingmar Weber. Demographic research with non-representative internet data. *International Journal of Manpower*, 36(1):13–25, 2015.

Joshua Wilde, Wei Chen, and Sophie Lohmann. Covid-19 and the future of us fertility: what can we learn from google? 2020.

Dawid Wladyka. The queries to google search as predictors of migration flows from latin america to spain, 2013.

Nebiye Yamak and Rahmi Yamak. Türkiye'de gelir dağılımı ve iç göç. *Dokuz Eylül Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, pages 16–28, 1999.

Dilek Yildiz, Jo Munson, Agnese Vitali, Ramine Tinati, and Jennifer A Holland. Using twitter data for demographic research. *Demographic Research*, 37:1477–1514, 2017.

Emilio Zagheni, Ingmar Weber, and Krishna Gummadi. Leveraging facebook's advertising platform to monitor stocks of migrants. *Population and Development Review*, pages 721–734, 2017a.

Emilio Zagheni, Ingmar Weber, and Krishna Gummadi. Leveraging facebook's advertising platform to monitor stocks of migrants. *Population and Development Review*, 43(4):721–734, 2017b.

Sinan Zeyneloglu, Ibrahim Sirkeci, Yaprak Civelek, et al. Language shift among kurds in turkey: A spatial and demographic analysis. *Kurdish Studies*, 4(1):25–50, 2016.