

Discussion Paper No. 18-008

**“Brain Gain” on Wikipedia:  
Immigrants Return Knowledge Home**

Olga Slivko

**ZEW**

Zentrum für Europäische  
Wirtschaftsforschung GmbH

Centre for European  
Economic Research

Discussion Paper No. 18-008

**“Brain Gain” on Wikipedia:  
Immigrants Return Knowledge Home**

Olga Slivko

Download this ZEW Discussion Paper from our ftp server:

**<http://ftp.zew.de/pub/zew-docs/dp/dp18008.pdf>**

Die Discussion Papers dienen einer möglichst schnellen Verbreitung von  
neueren Forschungsarbeiten des ZEW. Die Beiträge liegen in alleiniger Verantwortung  
der Autoren und stellen nicht notwendigerweise die Meinung des ZEW dar.

---

Discussion Papers are intended to make results of ZEW research promptly available to other  
economists in order to encourage discussion and suggestions for revisions. The authors are solely  
responsible for the contents which do not necessarily represent the opinion of the ZEW.

# “Brain Gain” on Wikipedia: Immigrants Return Knowledge Home\*

Olga Slivko\*\*

February 9, 2018

## Abstract

Economic literature acknowledges the impact of immigration on cross-border patenting and scientific publications. However, the role of immigration flows in the dissemination of knowledge in a broader sense is yet to be assessed. In this paper, I estimate the effect of immigration on the facilitation of online knowledge regarding destination countries in the native languages of immigrants. To quantify online knowledge, I focus on one of the world’s most viewed online knowledge platforms, Wikipedia.

I combine data on immigration flows between the pairs of immigrants’ origin and destination countries with contributions to Wikipedia describing the countries of immigrants’ destinations in the languages spoken in immigrants’ origin countries. I specifically focus on knowledge domains describing science and culture. In order to draw a causal inference, I use shocks to immigration due to economic and political crises as exogenous shocks to Wikipedia content and analyze subsequent changes in the contributions to Wikipedia.

An increase in immigration yields more knowledge contributed to Wikipedia about science and culture in destination countries in the native languages of the origin countries. Interestingly, the increase in contributions is driven by anonymous contributors. In the Wikipedia community, these are considered occasional contributors who care personally about the topics they contribute to. The increase in content generated anonymously is driven by longer contributions.

**Keywords:** Immigration; Knowledge dissemination; Online knowledge; Wikipedia.

**JEL Classification Numbers:** L17, O15, O33, O35, H41, L86.

---

\* I benefitted very much from the advice of Irene Bertschek, Shane Greenstein, Reinhold Kesler, Michael Kummer, Toomas Hinnosaar, Mary O’Mahony, Thomas Niebel, Marianne Saam, Patrick Schulte, Steffen Viete, Michael Ward, and Michael Zhang. I am also indebted to Tobias Werner, Niklas Isaak, Lukas Trotner and Martin Reinhard for their excellent research assistance. I would like to thank the participants of seminars at ZEW and IWM KMRC Tübingen. Financial support from Wissenschaftscampus Tübingen is gratefully acknowledged.

\*\*Address: L7, 1 68161 Mannheim, Telephone: +49 (0) 621 1235 358, E-mail: [slivko@zew.de](mailto:slivko@zew.de)

# 1 Introduction

The inflow of high-skilled immigrants fosters competitiveness and technological progress in destination countries by contributing to more dynamic knowledge creation and dissemination. An article published in *The Independent* in November, 2016 claims that “all six of America’s 2016 Nobel Prize winners are immigrants”\*. Economic studies widely support the important role played by immigration in the dissemination of knowledge, expressed, in particular, in the scientific publishing and patenting (Hunt and Gauthier-Loiselle (2010), Bosetti et al. (2015), Douglas (2015) and Ganguli (2015)). The present study addresses the dissemination of broader knowledge via Wikipedia, one of the world’s most viewed online knowledge repositories.

Although for the countries of immigrants’ origin the outflow of skilled individuals is often perceived as a negative phenomenon, leading to a deterioration of the knowledge economy and the quality of institutions in the origin countries (Anelli and Peri (2017), Docquier et al. (2016)), recent literature on innovation suggests that immigrants can act as a channel through which their origin countries could benefit from external knowledge to a certain extent. The positive impact of immigration on knowledge spillovers to the countries of immigrants’ origin has been shown for scientific publications and inventions. When researchers immigrate to the US, more patents from the US get cited by and cite patents from the countries of immigrants’ origin (Douglas (2015), Fackler et al. (2016)) and more scientific papers published in the US cite papers from the immigrants’ origin country (Ganguli (2015)).

Studies focusing on the spillovers to inventions and scientific publications might be capturing only the lower bound of the overall knowledge exchange taking place due to immigration. Moreover, in recent years, the availability of information and communication technologies, such as various online sites, and easier access to the Internet with tablets and smart phones, help individuals to become facilitators of knowledge. In addition, immigrants arriving from their origin countries might be better at expressing knowledge in such a way that it is more easily absorbed in their former homeland. These factors suggest that the amount of general knowledge disseminated due to immigration could go far beyond formal scientific knowledge. The dissemination of knowledge in a broader sense, i.e. meta-knowledge about science and culture, is the focus of this study. I analyze whether immigrants, who come to a destination country and learn about the local area, contribute information about the destination country to Wikipedia in their native languages.

I combine data on immigration flows to OECD countries from immigrant origin countries and contributions to Wikipedia covering country-specific knowledge domains in a set of languages. I exploit the fact that some knowledge domains, or, according to Wikipedia terminology, categories, such as, “Scientists” or “Research Institutes” focus on providing information in the category about each country, for example,

---

\*<http://www.independent.co.uk/news/world/americas/nobel-prize-winners-immigrants-us-donald-trump-brexit-immigration-racism-post-referendum-racism-a7355406.html>

“Research Institutes in Germany”. Information in these country-specific categories is represented in Wikipedia in multiple languages, including languages spoken in immigrant origin countries from OECD data. This allows me to match the yearly data on immigration flows between pairs of countries to yearly Wikipedia content added about the destination countries in the languages of immigrants’ origin countries. Then, to identify the effects of interest, I use the variation in the immigration flows and examine how it is related to the variation in content generation.\* Later, I additionally examine if the results hold in the natural experiment setting, in which countries experience economic and political crises representing shocks to immigration.

I show that, after immigration rises, more content contributions take place on Wikipedia in the knowledge-related domains, “Scientists” and “Research institutes” as well as in the culture-related domain “Cuisine” of each destination country. This content about scientists, research institutions and cuisine in destination countries is added to Wikipedia in the languages of origin countries. Moreover, for the knowledge-related domains more content is added mainly by anonymous contributors. The possibility to edit articles anonymously is an institutional feature of Wikipedia which aims at attracting more platform contributions from those users who do not care about building reputation in the community. Therefore, according to previous research on Wikipedia from computer science literature, immigration attracts occasional contributors to the online encyclopedia, who may have more expertise or be more interested in a specific topic as compared to loyal (registered) Wikipedia community members (Anthony et al. (2009)). Further exploration of anonymously contributed content suggests, that when we consider the location from where anonymous edits come, the increase in anonymous content in knowledge relevant domains is driven by anonymous contributions of one word and longer. These results are causal and they are robust to a variety of specifications.

Information contributed to Wikipedia has important implications for the technology adoption and individual knowledge-related choices (Thompson and Hanley (2017), Hinnosaar et al. (2017)). Therefore, considering immigration as another mechanism of knowledge dissemination can affect how policy makers approach immigration in both origin and destination countries. For example, contrary to the widely acknowledged hampering effect of outflow of human capital from the origin countries, having more knowledge about external science and research in the origin countries could promote the dissemination of new scientific methods and technologies and serve to forge contacts to external research bodies. In this respect, this paper relates to other studies that draw attention to how immigrants contribute to the economy in their origin countries through money remittances (Asatryan et al. (2017)), Foreign Direct Investment (Javorcik et al. (2011), Kugler and Rapoport (2007)), international trade (Gould (1994)) and international R&D collaborations (Miguelez (2016)).

---

\*Figure 9 shows an example of the variation between immigration flows and content generation that I explore. It shows anecdotal evidence that strong immigration from Spain to Germany after 2011 is related to intensive editing activity of an article in German Wikipedia about Max Planck Society in Spanish language. Similarly, immigration from Netherlands to Germany is low as is editing intensity of the article in Dutch.

## 1.1 Related Literature

This paper contributes to two strands of literature.

The first is the economic literature that aims to assess the knowledge dissemination effects of immigration. Studies highlight the importance of geographic proximity between researchers or inventors for acquiring relevant scientific knowledge and know-how (Ganguli (2015), Breschi and Lissoni (2009), Kerr (2008), Agrawal et al. (2011)). Since knowledge that can be codified and transmitted through scientific articles sometimes incorporates tacit knowledge that is easier to transmit in person, face-to-face interactions can be important for the generation of new ideas, as in the case of co-inventor networks in Breschi and Lissoni (2009). Kerr (2008) and Agrawal et al. (2011) show that ethnic scientific communities help to diffuse knowledge to immigrants' home countries in terms of patent citations. Ganguli (2015) suggests that pre-1990 Soviet scientific knowledge was spread to the US in the research fields in which Soviet scientists were active into the districts in the US where they moved after the collapse of the Soviet Union. However, Borjas and Doran (2015) point to a potential crowding-out effect on native workers by the arrival of skilled immigrants. Miguelez and Temgoua (n.d.) use a country-pair gravity model to explore knowledge feedbacks disseminated by inventors residing in host countries to their home countries. They find a 10% increase in the number of inventors of a given nationality to a destination country yields a 1% knowledge diffusion to their home country from the corresponding destination country. Finally, knowledge generated by the most technologically advanced countries flows farther (Peri (2005)).

Even after moving away from their home countries, immigrants may actively contribute to the well-being of their former co-patriots. Ghani et al. (2014) show that diaspora connections determine the choice of employees for outsourcing tasks: ethnic Indians are more likely to choose a worker in India when hiring on the online platform oDesk. This result highlights how diaspora continues to be important even in the online world, despite the efforts of oDesk to minimize frictions and provide full information about workers. On a related note, Agrawal et al. (2012) suggests that while workers in developing countries may face initial disadvantages on oDesk, these diaspora-based links could provide an opportunity to overcome initial uncertainty about workers. Such diaspora effects can be important for the economic integration of developing countries, their economic transition and growth. My study is similar in that it explores knowledge dissemination facilitated by an online platform, Wikipedia.

The second strand analyzes incentives of the crowd to engage time and effort to contribute online knowledge to Wikipedia, or, more generally, to participate in the provision of online public goods. Studies show that social spillovers and network structure promote content contributions to Wikipedia (Zhang and Zhu (2011), Piskorski and Gorbatai (2013), Hergueux et al. (2014), Kummer (2013) and Slivko (2014)), while this study indicates that an additional channel, learning or gathering new information, can play a role in content generation on Wikipedia. However, the effects that I find in this study could be reinforced by social spillovers or themselves create social spillovers on Wikipedia.

Online knowledge freely available on Wikipedia can have very salient economic implications, as shown in field experiments by Hinnosaar et al. (2017) and Thompson and Hanley (2017). Hinnosaar et al. (2017) show that the amount of content about sightseeing in Spanish towns attracts more tourists to these towns, while Thompson and Hanley (2017) show that adding new scientific content to Wikipedia leads to the usage of words added in the scientific literature. These are the first studies to show that knowledge in the online repository contributes to decision-making by individuals and affects technological progress, which highlights the importance of meta-knowledge disseminated by immigrants.

The remainder of the paper is structured as follows. Section 2 describes the data. Section 3 presents the empirical analysis and reports the results on the effect of immigration on knowledge dissemination. Section 4 analyzes the nature of anonymous contributions and section 5 takes into account the location of contributions. Finally, Section 6 concludes.

## 2 Data

### 2.1 Migration

To analyse the effects of immigration on knowledge diffusion I combine data from several sources. For immigration flows, OECD provides publicly available International Migration Database. It contains information on the inflow, outflow and stock of migrants for each country of destination belonging OECD distinguishing the migrants' countries of origin. The data covers the years 2000 to 2015.

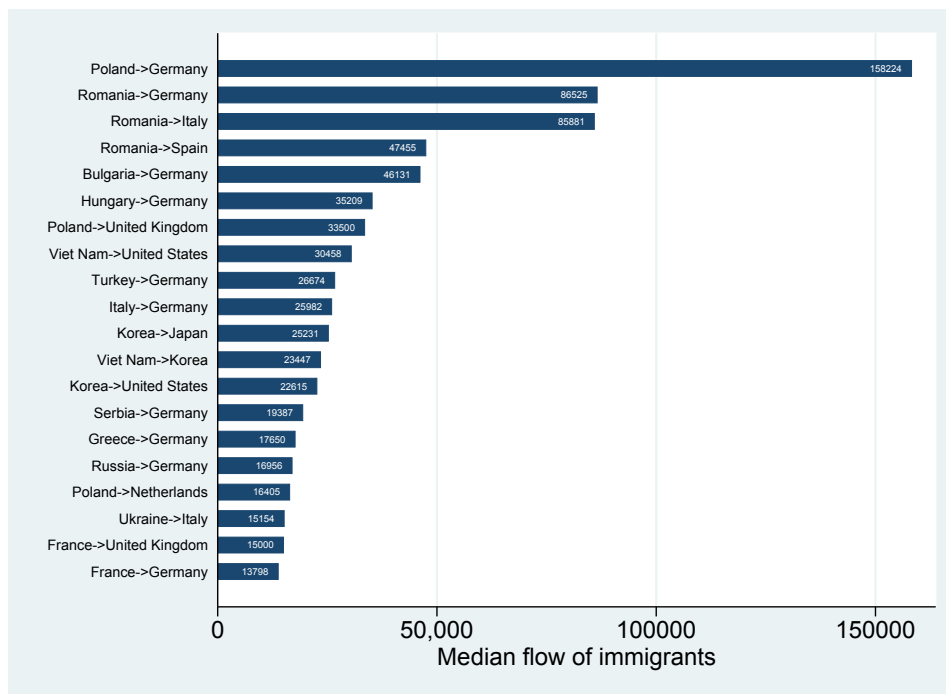
OECD yearly data on migration do not provide information on the education levels of migrants. To distinguish between immigrants with different skill levels in my analysis, I add the data from the Database on Immigrants in OECD Countries (DIOC). The DIOC gathers data from the survey which was conducted in years 2000/01, 2005/06 and 2010/11. The questions in the survey cover the age and nationality, duration of stay and labour force status of respondents. I use educational levels of foreign-born individuals who at the time of the survey had not yet been granted the citizenship in their host country. Based on this number, I construct the weights of individuals with tertiary education (education following the completion of a school, including universities as well as trade schools and colleges) between each pair of origin and destination countries. Then I apply these shares to OECD yearly immigration flows. Since data on the shares of education levels are only available for the years 2006 and 2011, I impute the missing shares using linear trends between the years 2006 and 2011, and for the remaining years in the sample after 2011, I use shares of 2011. Since these shares are available for only a limited number of countries, all estimations which use migrants' education levels employ fewer observations.

Furthermore, I merge data on immigration flows with the measures of knowledge dissemination from Wikipedia. Figure 1 presents the median flow of immigrants between the pair of origin and destination countries for the pairs with the highest immigration (in the resulting sample). As we will see, high levels

of immigration are observed between the neighbouring countries, Germany and Poland, and, in general, Germany is a very popular destination country. The bars of the histogram also indicate high levels of immigration not only to Europe, but also between Asian countries and between European and North American countries. Furthermore, in the empirical analysis I check the robustness of the results when the largest destination and origin countries are excluded from the sample.

The shares of immigrants with tertiary education (university degree) for the top 20 country pairs for which data are available are presented in Figure 2. The shares equal to 1 for the origin countries Bulgaria and France and the destination country Germany are due to the fact that the DIOC data do not contain information about other education groups than for the tertiary education for these origin countries.

Figure 1: Top 20 immigration inflows to OECD countries in 2006-2015.

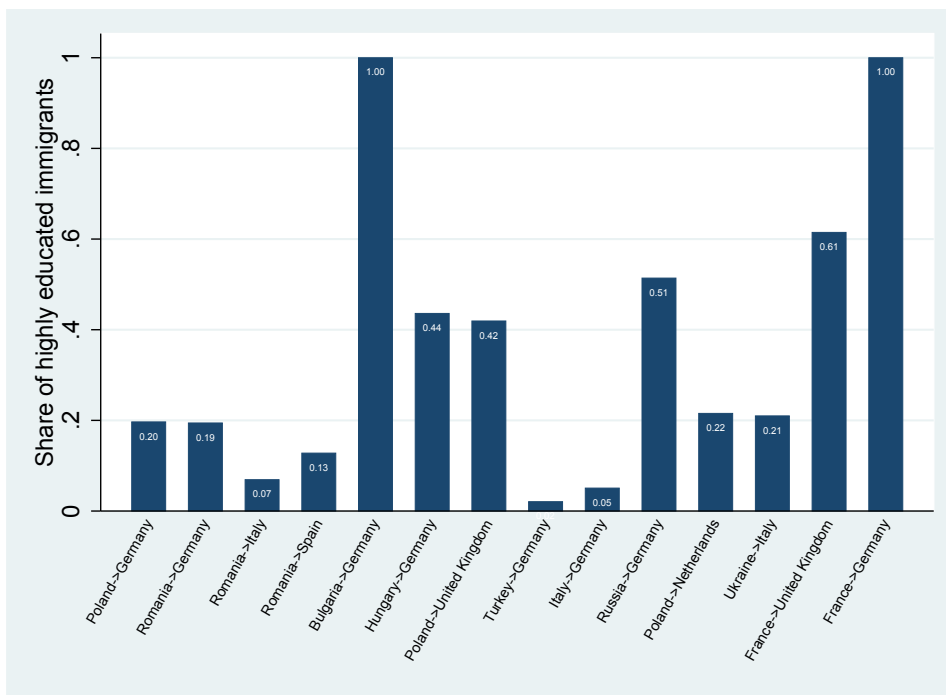


NOTE: This figure illustrates the pairs of countries with the highest median immigrant inflows.

Immigration in the resulting sample is not driven primarily by the flows from developing to developed countries. Figure 3 displays total outflow of immigrants from different groups of countries (OECD, EU, non-OECD, non-EU countries) in each year. Before 2009, immigration flows from all groups of countries were to some extent declining, while after 2010 there was some growth, especially in immigration from European countries, potentially driven by the economic crises in Europe which began in 2008. After 2013, immigration from non-OECD and non-EU countries grows, but by less than 0.5 standard deviations.

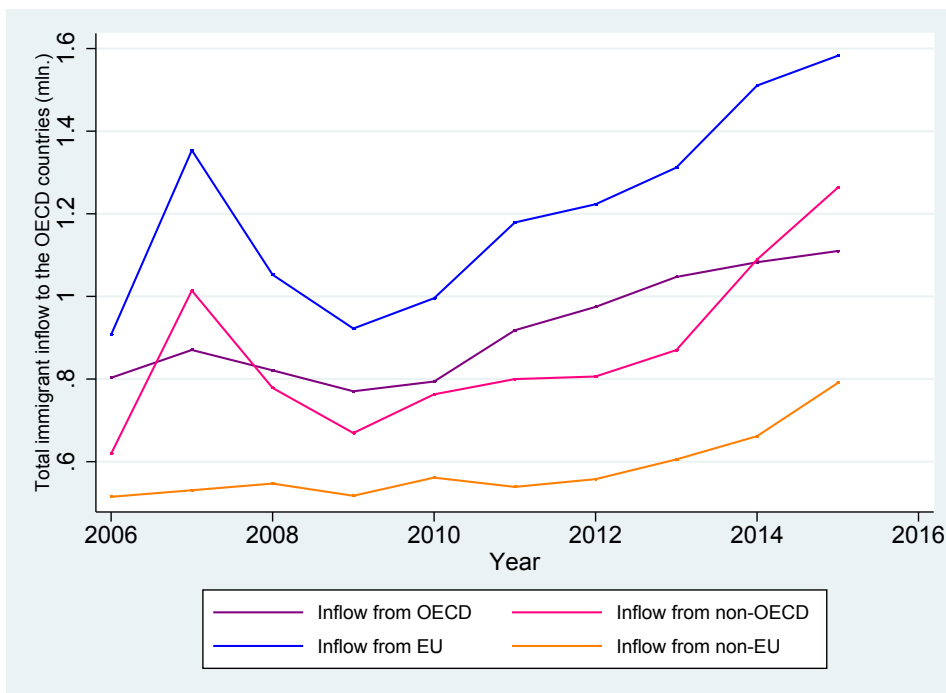


Figure 2: The shares of immigrants with tertiary education for top 20 immigration inflows to OECD countries in 2006-2015.



NOTE: This figure illustrates for each pair of countries from Figure 1 the share of immigrants with tertiary education based on DIOC OECD data.

Figure 3: Immigration inflows to OECD countries in 2006-2015.



NOTE: This figure illustrates time trends for the total number of immigrants entering OECD countries from OECD, non-OECD, European and non-European countries.

## 2.2 Background on knowledge domains on Wikipedia

As the largest and the most viewed free online encyclopedia, Wikipedia represents an ideal setting for measuring the dissemination of online knowledge.\* While Wikipedia articles cover a huge number of topics, I focus on online knowledge in several domains. To measure the relevant meta-knowledge describing science and technology that could potentially be disseminated by migrants, I focus on domains “Research institutes”, which describes research institutions, universities and academic societies, and “Scientists”, which lists the names of scientists. In addition, to be able to compare meta-knowledge relevant for research and science with broader knowledge accessed by the general public, I add into the analysis the culture-related domain “Cuisine”.

Wikipedia has a specific feature which makes it possible to combine content contributed with data on immigration flows between countries. Most articles in Wikipedia belong to knowledge domains, or categories. Wikipedia community members create categories and label articles with related categories. Among these categories, there are a few which classify knowledge by topic and by country, for example, “Research Institutes” in each country. Figure 4a presents the articles of the category “Research institutes in France” in the English language. Hence, these country-specific categories can represent the available knowledge about the destination countries of migrants. These country-specific categories exist on Wikipedia in many languages. This allows me to match knowledge available about destination countries in the languages of the migrants’ origin countries with the origin-destination immigration flows from OECD.

I extract information from the selected domains on Wikipedia using an API tool in the following way. I determine the set of main categories describing research institutions of each country in English, for example, “Research institutes in France” (Figure 4). For each category I download language links, as seen in the left-lower corner in Figure 4a. These language links tell me in which languages I can obtain information about each main category, describing knowledge about the destination countries of migrants. Then, within each language listed in the language links I can extract subcategories of the main category and articles that are directly related to the category (see upper and lower panel in Figure 4a). Subsequently, for each subcategory I can go one level deeper in the categorical tree. For example, Figure 4b shows the outcome of going one level deeper for the subcategory “Computer science institutes in France”. For each subcategory, I collect subcategories and articles belonging to it and so on. My goal is to automatically collect as many articles belonging to each domain as possible, while avoiding the inclusion of articles that do not directly correspond to the domain. Therefore, for the domain “Research institutes” I collected articles belonging to the main category and to the subcategories of the first level, which means in two iterations. For the domain “Scientists” it took me three iterations. At each step I

---

\*Wikipedia is the 5th most visited site in World Wide Web, after Google, Youtube, Facebook and Baidu according to alexa.com (October, 2017).

Category: Research institutes in France

From Wikipedia, the free encyclopedia

**Subcategories**

This category has the following 6 subcategories, out of 6 total.

- C**
  - Computer science institutes in France (9 P)
- F**
  - French National Centre for Scientific Research (2 C, 90 P)
- G**
  - Grenoble Institute of Technology (3 C, 14 P)
- P**
  - Paris Institute of Criminology people (1 C)
  - University of Paris-Saclay (1 C, 5 P)
  - Pasteur Institute (26 P)

**Pages in category "Research institutes in France"**

The following 102 pages are in this category, out of 102 total. This list may not reflect recent changes (learn more).

- A**
  - Aix-Marseille University Faculty of Sciences
  - Asia Centre
  - Astroparticle and Cosmology Laboratory
- C**
  - Cadarache
  - Catholic Office of Information and Initiative for Europe
  - CCMIX
  - Center for Research and Restoration of Museums of France
- G**
  - Grenoble Institute of Technology
- H**
  - Higher Institute of Iranian Studies
- I**
  - IATEUR
  - IBM La Gaude
  - IFREMER
  - IFSTTAR
  - Institut Charles Sadron
- L**
  - Laboratoire atmosphères, milieux, observations spatiales
  - Laboratoire d'ethnologie et de sociologie comparative
  - Laboratoire d'informatique de Paris 6
  - Laboratoire d'informatique pour la mécanique et les sciences de l'ingénieur
  - Laboratoire de droit des affaires et nouvelles technologies
  - Laboratoire de Physique des Solides
  - Laboratoire de Zététique
  - Laboratoire des sciences du climat et de l'environnement
  - Laboratoire Plasma et Conversion d'Énergie
  - Laboratory for Analysis and Architecture of Systems
- N**
  - National Computer Center for Higher Education (France)
- O**
  - ONERA

(a) Main category “Research institutes in France” in English (with its subcategories and articles)

Category: Computer science institutes in France

From Wikipedia, the free encyclopedia

**Pages in category "Computer science institutes in France"**

The following 9 pages are in this category, out of 9 total. This list may not reflect recent changes (learn more).

- F**
  - French Institute for Research in Computer Science and Automation
- H**
  - Hubert Curien Laboratory
- I**
  -
- L**
  - Laboratoire d'Automatique, Génie Informatique et Signal
  - Laboratoire d'Informatique de Grenoble
  - Laboratoire d'Informatique de Paris 6
  - Laboratoire d'Informatique Fondamentale de Lille
- N**
  - National Computer Center for Higher Education (France)
- O**
  - ONERA

Categories: [Computer science institutes by country](#) | [Research institutes in France](#) | [Computer science in France](#)

(b) A subcategory of main category “Computer science institutes in France” (with its subcategories and articles)

Figure 4: Main category and its subcategory on Wikipedia: Research institutes of France in English language

was comparing the fit of the articles I gain by going one step further and the amount of articles that do not fit and appear in the sample. For the domain “Cuisine” in a country, it took five iterations to collect articles that fit the domain well.

After collecting all articles that describe knowledge domains in various languages, I extract the full revision history of each article. This records all edits for each article, and for each edit the information available includes ID, time, the name of the user who made the edit, the size of the article in bytes after

the edit. Therefore, various measures of content can be derived from the article revision history. For each article, I compute the number of edits and the amount of bytes contributed, distinguishing between those made by registered and anonymous users.\* After that, I aggregate the measures of content contributions at the level of knowledge domain and language, for example, the amount of content contributed yearly about “Research institutes in Germany” in French. I collect the revision history of articles dating back to 2006, because this was the year that Wikipedia became the most popular reference website on the Internet, according to Hitwise.\*

Merging data on immigration flows with indicators of online content generation on Wikipedia between the years 2006 and 2016 yields the final data set with observations on immigration and online content generation in the period of 2006-2015. In order to properly merge Wikipedia content about destination countries in the languages of origin countries with immigration flows, I have to exclude origin countries with potentially high levels of illegal immigration (some Middle Eastern countries, such as Afghanistan, Lybia and Syria). Including these countries would introduce a measurement error into the data. I also exclude countries whose main languages are spoken all over the world, such as English, Spanish, German, Portuguese, Persian or Arabic, because it would difficult to attribute knowledge about a destination country contributed in these languages, for example, in English, to any particular country of migrants’ origin. Importantly, I exclude China from my dataset, because Wikipedia has been banned there since 2008. As Zhang and Zhu (2011) show, following the drastic reduction in contributions from mainland China, contributions in Mandarin from other neighbouring Mandarin-speaking countries, such as Singapore or Taiwan, also decreased. The authors attribute such a reaction to the decrease in the size of the recipient group. As for individuals who immigrate from China, since Wikipedia was not an important online information site while they still lived in their home country, it is unlikely that they would decide to contribute to it after they moved abroad.

The final set of origin countries, in whose languages there were contributions in at least one of the three Wikipedia domains, comprises about 39 origin countries in Europe and Asia (see also Table 1) and 34 host countries (see Table 2). In the resulting sample, Figure 6 (see Appendix) shows the strongest knowledge flows in the domain “Scientists”, Figure 7 shows which Wikipedia content flows between origin and destination countries have been the strongest in the domain “Research Institutes”, and Figure 8 shows the most relevant content flows for the domain “Cuisine”. As expected, in “Research Institutes”, the countries which receive the most content in the languages of immigrants are United States, German and France. Similarly, for “Scientists”, the countries with the most contributions are Germany, the UK and the US, all global leaders in scientific research. For “Cuisine” the composition of countries is more varied.

---

\*Wikipedia allows users to edit articles after the log-in process as well as directly, without logging-in. In the former case, the data contain the username of the user who made an edit, and, in the latter case, only the IP address of the contributor.

\*[https://en.wikipedia.org/wiki/History\\_of\\_Wikipedia](https://en.wikipedia.org/wiki/History_of_Wikipedia)

The Figure 10 (see Appendix) shows how anonymous and registered edits about “Scientists” vary over time. On average, edits decrease as the coverage of topics grows with time and tends to saturate.

### 2.3 Descriptive statistics

After merging the OECD immigration data with Wikipedia content generation indicators, I obtain the sample composed of the following origin countries (Table 1) and destination countries (Table 2).

The countries of immigrants’ origin include European countries, former Soviet Union countries and Asian countries (Table 1). The highest outflows of migrants could be observed from Poland and Romania. The destination countries include all OECD countries (Table 2). The table shows that strong immigration inflows into countries like the US, Germany, or Italy are associated with large amounts of content added to the research-related domains, scientists and institutes, as well as cuisine of the country.

Table (3) displays descriptive statistics for the explanatory variables of the analysis. On average, 2,955 immigrants in my sample leave their country of origin and move to a new destination country. Among them, the share of immigrants with tertiary education is on average 0.4, which gives on average 838 immigrants with tertiary education.

Considering the content of Wikipedia editions in native languages of immigrants, the average language edition sees a yearly increase in the number of registered Wikipedia users by 2,527 persons, and receives 179 new articles and 2,051 edits each year. Contributors create 0.2 new articles per day.

Table 4 shows summary statistics for knowledge domains indicators, which I use as dependent variables throughout this study. The average yearly edits for the domain “Research institutes” (16.9) is on average lower than for the domain “Cuisine” (93.6). The domain “Scientists” gets on average 373.3 edits per year. The measures of Wikipedia content, as well as of immigration, appear to be highly skewed, so in the estimations I use their natural logarithms.

Table 1: Immigration inflows from the source countries over the years 2007-2015.

	Total Emigrants(K)	Av. Yearly Emigrants(K)	Std. dev.	Maximum
Albania	144.96	7.25	14.99	69.36
Armenia	75.87	0.95	1.28	6.32
Azerbaijan	36.20	0.26	0.47	3.15
Belarus	84.18	0.58	0.84	3.18
Bulgaria	882.35	3.02	9.79	86.27
Croatia	283.30	1.73	6.50	60.98
Czech Republic	185.37	0.71	1.64	10.97
Denmark	119.60	0.52	0.91	5.14
Estonia	81.75	0.38	0.89	6.04
Finland	105.98	0.43	0.63	2.81
France	823.31	2.85	4.13	24.00
Georgia	65.02	0.96	1.36	6.47
Greece	309.54	1.39	4.54	32.66
Hungary	645.80	2.78	8.76	59.99
Iceland	10.10	0.15	0.26	1.17
Ireland	12.86	1.61	0.14	1.91
Israel	115.17	0.52	1.06	5.94
Italy	973.29	3.36	7.54	57.19
Japan	350.76	1.35	1.93	8.27
Kazakhstan	35.43	0.61	0.72	2.56
Korea	730.88	2.80	6.38	30.04
Latvia	135.28	0.84	1.85	10.03
Lithuania	213.19	1.39	2.88	17.00
Macedonia	131.48	1.01	3.07	24.78
Mongolia	16.58	0.40	0.29	1.30
Netherlands	370.14	1.25	2.54	11.70
Norway	90.74	0.53	0.75	4.00
Poland	2815.47	9.68	29.60	192.17
Romania	3271.61	18.48	42.14	271.44
Russia	747.94	2.62	3.91	31.37
Serbia	367.11	2.04	5.25	39.72
Slovak Republic	318.63	1.39	2.88	15.52
Slovenia	58.54	0.28	0.73	4.75
Sweden	195.21	0.71	1.17	8.20
Thailand	499.65	3.73	5.97	48.33
Turkey	597.63	2.44	5.20	29.59
Ukraine	843.91	3.16	5.91	45.24
Viet Nam	938.28	4.86	10.01	65.85

NOTE: In columns (1)-(4), this table shows main statistic measures of aggregate immigration flows from each source country: the total emigration in the period of observation, average yearly outflow from the origin country per country of emigrants' destination, standard deviation and maximum values. All values are computed only for the sample used in the estimations.

Table 2: Total immigration inflows into host countries and total number of edits in the categories and about each host country over the years 2007-2015.

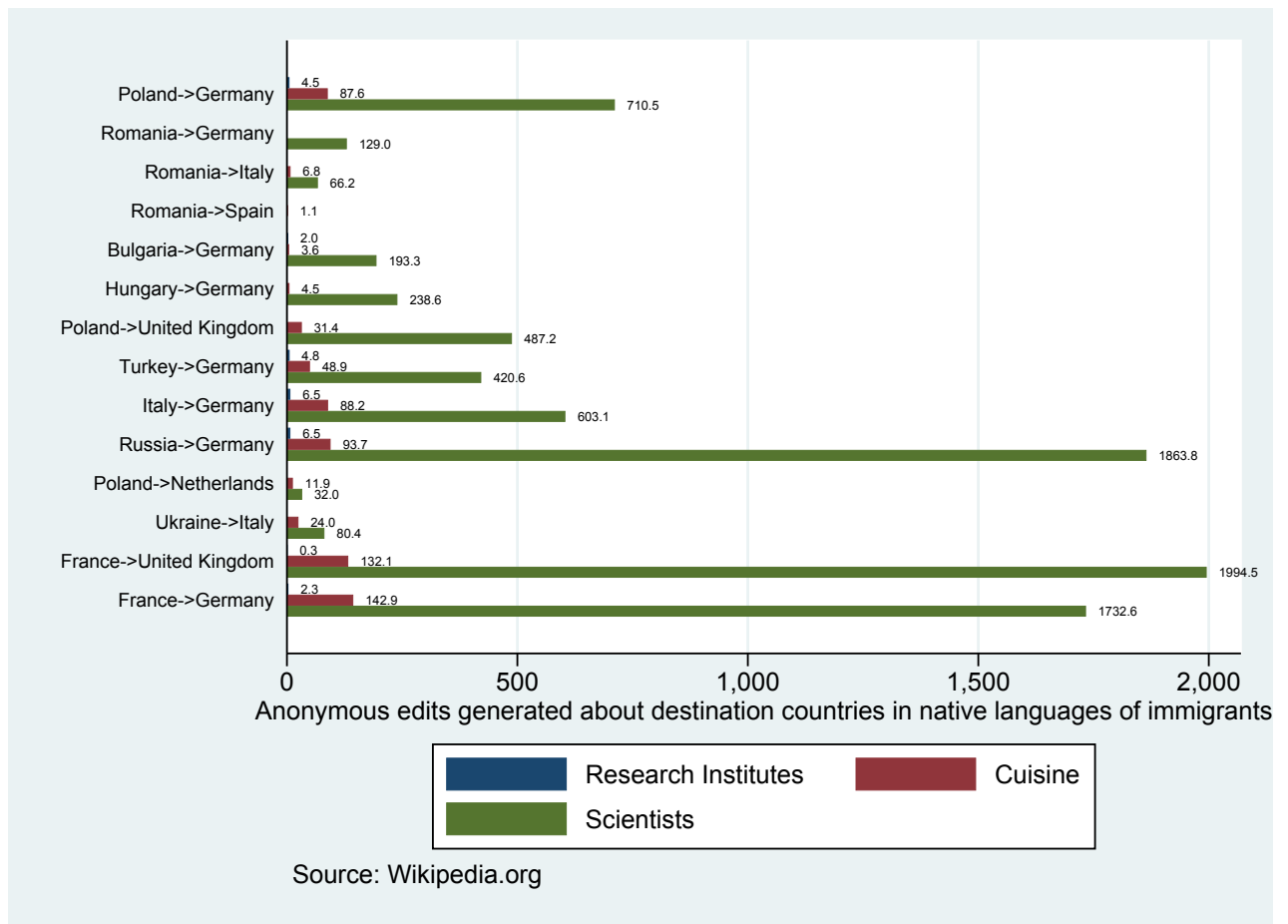
	Total Immig.(K)	Skilled Immig.(K)	Total Edits, <i>Res.Institutes</i>	Total Edits, <i>Cuisine</i>	Total Edits, <i>Scientists</i>
Australia	180.40	46.48	126	3123	18014
Austria	655.66	.	9	9794	103112
Belgium	458.72	40.51	583	12374	78830
Canada	269.35	109.11	344	10171	73476
Chile	20.10	4.94	0	1814	4640
Czech Republic	296.12	78.89	0	5847	26960
Denmark	183.12	70.40	90	3858	42205
Estonia	14.52	3.50	65	1802	8345
Finland	102.10	13.42	109	4759	26862
France	234.74	53.42	1656	53695	384132
Germany	6040.63	997.51	2514	28862	666356
Greece	13.31	2.62	0	2068	11514
Hungary	137.95	13.55	0	7147	40659
Iceland	27.46	2.10	360	1054	2539
Israel	32.54	.	0	7968	13694
Italy	1508.71	110.67	154	55159	180821
Japan	793.71	112.32	228	43608	32305
Korea	516.02	.	0	20676	5138
Latvia	13.01	.	390	1262	4968
Luxembourg	37.71	16.28	0	183	711
Mexico	25.41	7.12	0	13834	5549
Netherlands	476.98	80.23	56	5985	65562
New Zealand	88.66	.	9	1282	5376
Norway	292.37	96.90	90	3053	28590
Poland	272.79	70.84	234	11330	77035
Portugal	53.59	6.61	0	3108	4396
Slovak Republic	50.02	2.63	0	4212	8110
Slovenia	78.32	6.17	0	774	3630
Spain	1355.66	205.27	30	19971	32174
Sweden	322.21	107.12	321	5925	68930
Switzerland	476.64	4.28	387	11187	108331
United Kingdom	1203.00	489.89	157	16401	263092
United States	1451.60	392.28	5544	49015	831658

NOTE: For each destination country columns (1)-(5) display the total number of immigrants and the number of skilled immigrants coming to each country of destination (in rows) together with the total number of edits in each observed domain aggregated across all countries of immigrants' origin.

Table 3: Descriptive statistics of immigration flows

	Mean	Std. dev.	Min	Max	# Obs.
Immigrants	2450	10487	0	271443	7217
Share of high-skilled immigrants	.408	.173	0	1	3754
High-skilled immigr.	838	3636	0	86274	3754
$\Delta$ Immigrants, k	.116	4.31	-136	232	6514
New Wikipedians in language	2188	2785	14	12106	7177
Total articles in language, k	352	420	2.3	2400	7177
New articles per day in language, k	.159	.255	.00175	2.85	7177
All edits in language, k	1786	2091	6.34	9595	7177

Figure 5: Wikipedia content contributed by anonymous users for country pairs from top 20 by immigration inflows in 2006-2015.



NOTE: This figure illustrates how many edits per year on average are made about destination countries in native languages of origin countries for 20 country pairs with the highest median immigration flows in the sample.



Table 4: Descriptive statistics of Wikipedia content

	Mean	SD	Min	P25	P50	P75	Max	Count
Scientists, kb	184.9	541.0	0.0	4	23	117	8293.7	4233
Scientists, an.kb	45.5	165.3	0.0	0	2	19	3048.2	4233
Scientists, reg.kb	139.4	396.1	0.0	3	19	93	5948.8	4233
Scientists, edits	705.3	1879.9	0.0	34	129	508	24336.0	4233
Scientists, an.edits	116.0	340.3	0.0	3	16	70	4116.0	4233
Scientists, reg.edits	589.3	1563.2	0.0	30	108	418	20990.0	4233
Scientists, an.minor edits	44.5	130.4	0.0	1	6	28	1688.0	4198
Scientists, an.larger edits	63.1	192.8	0.0	1	8	37	2350.0	4198
Research instit., kb	2.8	6.3	0.0	0	0	2	57.2	795
Research instit., an.kb	0.6	2.2	0.0	0	0	0	29.0	795
Research instit., reg.kb	2.2	5.1	0.0	0	0	2	46.1	795
Research instit., edits	16.8	24.1	0.0	2	8	21	187.0	795
Research instit., an.edits	2.5	5.5	0.0	0	1	3	63.0	795
Research instit., reg.edits	14.3	20.4	0.0	2	7	18	159.0	795
Research instit., an.minor edits	0.7	1.9	0.0	0	0	1	23.0	758
Research instit., an.larger edits	1.5	3.4	0.0	0	0	1	37.0	758
Cuisine, kb	0.1	0.2	0.0	0	0	0	3.0	4778
Cuisine, an.kb	4.7	27.2	0.0	0	0	1	1155.6	4778
Cuisine, reg.kb	10.8	35.4	0.0	0	1	6	680.0	4778
Cuisine, edits	88.1	196.8	0.0	7	24	75	3149.0	4778
Cuisine, an.edits	22.2	58.3	0.0	1	4	17	1179.0	4778
Cuisine, reg.edits	66.0	147.6	0.0	5	19	58	3119.0	4778
Cuisine, an.minor edits	9.7	25.1	0.0	0	2	8	563.0	4670
Cuisine, an.larger edits	11.2	31.9	0.0	0	2	7	605.0	4670

Note: Unit of observation is Wikipedia content about a destination country in the language of origin country over a year.

### 3 Immigration flows and Wikipedia

#### 3.1 Baseline approach

To estimate the impact of the immigration flows on knowledge dissemination, I exploit the variation between the inflow of immigrants from a set of origin countries and the amount of knowledge generated about host countries in the immigrants' native languages. For that, I regress the amount of content generated on Wikipedia about host countries in the immigrants' native languages on the independent variables of interest. As an independent variable, depending on a specification, I use the log number of migrants, the log number of high-skilled immigrants and the increase in the absolute number of migrants. This yields the following regression equation:

$$Y_{dot} = \alpha_{do} + \alpha_t + \beta Immigration_{dot} + X_{ot} \gamma + X_{dt} \delta + \epsilon_{dot}, \quad (1)$$

where  $d$  stands for the country of destination (for immigration flows) or the topic dedicated to the destination country on Wikipedia,  $o$  points to the country of immigrants' origin or the language of the content,  $t$  is the year of observation.  $Y_{dot}$  is the amount of content about destination countries in immigrants' native languages, i.e. the languages commonly used in their origin countries. Country pair fixed effects,  $\alpha_{do}$ , are included to control for the time-invariant heterogeneity between the pair of countries, for example, for the popular migration destinations for every origin country, which might drive online content presence. For example, if Germany is a popular destination country for migrants from Turkey due to diaspora connections, there would be more content about Germany available in Turkish.  $\alpha_t$  are year fixed effects to control for common trends in online content generation from year to year.

To measure  $Immigration_{dot}$ , I use three indicators and present them together in the tables with results. First, I use the logarithm of the total number of immigrants between the country of origin and the country of destination. Second, I use the measure of skilled immigration between countries. For that I multiply the total number of immigrants by the weight of immigrants with tertiary education from the same origin in the stock of migrants in each destination country. Third, I approximate the shock to immigration by computing the absolute change in the number of immigrants between two countries in the current year as compared to the previous year. Then, the fixed effects within estimation would capture the change of the dependent variable in response to the change in the pace at which immigration between the pair of countries grows or decreases.

In the robustness check, I use a set of control variables,  $X_{ot}$ . Some language editions of Wikipedia may grow faster than others, and this could affect the increase in the amount of information in a particular domain in the origin language of immigrants just because of more general growth of content in this language. To control for language specific content growth, I check the robustness adding as a control variable the number of edits in Wikipedia language editions in each year  $t$ , where language editions

correspond to the countries of immigrant origin  $o$ . The dynamics of this control variable captures the development of every language edition of Wikipedia over time due to other reasons than immigration (for instance, social spillovers on the platform).

Similarly, to account for growth in the domains describing destination countries, I add as a control variable,  $X_{dt}$ , the amount of content generated about the destination country in English or in the native language of destination countries,  $d$ . This helps to tease out the variation in the content about the destination country. For example, there might be more content about scientists in Germany and in the US just because of recent advances in science. Controlling for the amount of content in English language would account for this variation.

In all estimations, I cluster standard errors by origin-destination country pair to allow for serial correlations in the immigration flows between origin and destination countries. In the robustness check, I show that the effects I find are robust to clustering standard errors at the level of origin countries.

## 3.2 Results

The estimation results for equation (1) are presented in Tables 5, 6 and 7. In all tables, the independent variables of interest are (i) the logarithm of the number of immigrants (columns (1)-(4)), (ii) high-skilled immigrants (columns (5)-(8)) and (iii) the absolute increase in the number of immigrants (columns (9)-(12)). For each knowledge domain, I have the following dependent variables: the number of edits made by anonymous (1) and registered (2) users, and the amount of bytes contributed by anonymous (3) and registered (4) users. All dependent variables are in natural logarithms (plus one, to treat zero values).

Table 5: Immigration Flows and Content Generation on Wikipedia in domain *Scientists*: Fixed Effects Estimation.

	Total Immigrants				Higher Education				$\Delta$ Immigrants, k			
	(1) Anonym. Edits	(2) Reg. Edits	(3) Anonym. Bytes	(4) Reg. Bytes	(5) Anonym. Edits	(6) Reg. Edits	(7) Anonym. Bytes	(8) Reg. Bytes	(9) Anonym. Edits	(10) Reg. Edits	(11) Anonym. Bytes	(12) Reg. Bytes
Lg Immigrants	-0.023 (0.017)	-0.038* (0.021)	0.046 (0.058)	-0.068 (0.049)								
Lg Skilled Immigrants					-0.009 (0.033)	-0.055 (0.040)	0.098 (0.096)	-0.070 (0.082)				
$\Delta$ Immigrants, k									0.002*** (0.001)	-0.000 (0.001)	0.003 (0.002)	0.001 (0.001)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Host country time trends	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean dep. Variable	2.03	3.84	5.05	8.27	2.10	3.94	5.23	8.43	1.99	3.81	5.00	8.24
Observations	5854	5854	5854	5854	3176	3176	3176	3176	5287	5287	5287	5287
# of country pairs	690	690	690	690	486	486	486	486	673	673	673	673
R <sup>2</sup>	0.092	0.086	0.060	0.054	0.078	0.072	0.072	0.059	0.084	0.096	0.066	0.065

NOTE: This table contains estimation results for different measures of Wikipedia content (the amount of bytes contributed by anonymous and registered users, the number of edits by anonymous and registered users, minor edits and new articles) about host countries on the languages of origin countries. Columns (1) - (4) show the results where the independent variable of interest is log number of immigrants from origin country  $o$  to destination country  $d$  in year  $t$ . In columns (5) - (8) the independent variable of interest is the log number of immigrants with tertiary education. Finally, in columns (9) - (12) I use as the independent variable an absolute increase in the number of immigrants in year  $t$ , as compared to year  $t-1$ . All specifications include year dummies. All standard errors (in parentheses) are clustered at the origin-destination country pair level. Significance stars denote: \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ .

Table 6: Immigration Flows and Content Generation on Wikipedia in domain *Research Institutes*: Fixed Effects Estimation.

	Total Immigrants				Higher Education				$\Delta$ Immigrants, k			
	(1) Anonym. Edits	(2) Reg. Edits	(3) Anonym. Bytes	(4) Reg. Bytes	(5) Anonym. Edits	(6) Reg. Edits	(7) Anonym. Bytes	(8) Reg. Bytes	(9) Anonym. Edits	(10) Reg. Edits	(11) Anonym. Bytes	(12) Reg. Bytes
Lg Immigrants	0.046** (0.021)	0.039 (0.029)	0.266*** (0.094)	0.098 (0.090)								
Lg Skilled Immigrants					0.051* (0.026)	0.016 (0.026)	0.259** (0.124)	0.082 (0.107)				
$\Delta$ Immigrants, k									-0.000 (0.010)	0.005 (0.006)	0.061* (0.036)	0.005 (0.015)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Host country time trends	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean dep. Variable	0.71	2.03	2.22	5.23	0.81	2.16	2.58	5.43	0.72	2.04	2.27	5.22
Observations	822	822	822	822	494	494	494	494	749	749	749	749
# of country pairs	127	127	127	127	82	82	82	82	121	121	121	121
R <sup>2</sup>	0.053	0.093	0.090	0.140	0.051	0.133	0.115	0.204	0.040	0.111	0.076	0.159

NOTE: This table contains estimation results for different measures of Wikipedia content (the amount of bytes contributed by anonymous and registered users, the number of edits by anonymous and registered users, minor edits and new articles) about host countries on the languages of origin countries. Columns (1) - (4) show the results where the independent variable of interest is log number of immigrants from origin country  $o$  to destination country  $d$  in year  $t$ . In columns (5) - (8) the independent variable of interest is the log number of immigrants with tertiary education. Finally, in columns (9) - (12) I use as the independent variable an absolute increase in the number of immigrants in year  $t$ , as compared to year  $t-1$ . All specifications include year dummies. All standard errors (in parentheses) are clustered at the origin-destination country pair level. Significance stars denote: \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ .

Table 7: Immigration Flows and Content Generation on Wikipedia in domain *Cuisine*: Fixed Effects Estimation.

	Total Immigrants				Higher Education				$\Delta$ Immigrants, k			
	(1) Anonym. Edits	(2) Reg. Edits	(3) Anonym. Bytes	(4) Reg. Bytes	(5) Anonym. Edits	(6) Reg. Edits	(7) Anonym. Bytes	(8) Reg. Bytes	(9) Anonym. Edits	(10) Reg. Edits	(11) Anonym. Bytes	(12) Reg. Bytes
Lg Immigrants	0.062** (0.025)	0.078*** (0.029)	0.128* (0.075)	0.115 (0.079)								
Lg Skilled Immigrants					0.089*** (0.034)	0.105** (0.048)	0.159 (0.100)	0.106 (0.105)				
$\Delta$ Immigrants, k									0.006*** (0.001)	0.003*** (0.001)	0.009 (0.007)	0.005 (0.004)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Host country time trends	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean dep. Variable	1.81	3.00	4.20	6.45	1.91	3.07	4.38	6.59	1.80	2.99	4.15	6.41
Observations	4867	4867	4867	4867	2526	2526	2526	2526	4435	4435	4435	4435
# of country pairs	642	642	642	642	421	421	421	421	608	608	608	608
R <sup>2</sup>	0.051	0.032	0.048	0.039	0.042	0.039	0.068	0.061	0.040	0.022	0.052	0.041

NOTE: This table contains estimation results for different measures of Wikipedia content (the amount of bytes contributed by anonymous and registered users, the number of edits by anonymous and registered users, minor edits and new articles) about host countries on the languages of origin countries. Columns (1) - (4) show the results where the independent variable of interest is log number of immigrants from origin country  $o$  to destination country  $d$  in year  $t$ . In columns (5) - (8) the independent variable of interest is the log number of immigrants with tertiary education. Finally, in columns (9) - (12) I use as the independent variable an absolute increase in the number of immigrants in year  $t$ , as compared to year  $t-1$ . All specifications include year dummies. All standard errors (in parentheses) are clustered at the origin-destination country pair level. Significance stars denote: \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ .

The results for the knowledge relevant domain “Scientists” of destination countries are presented in Table 5. They suggest that an increase in the number of immigrants positively affects content generation on Wikipedia about destination countries in the languages spoken in the countries of immigrant origin by anonymous users (column (9)). The coefficient of 0.002 can be interpreted as follows: a one unit increase in the flow of immigrants from country  $o$  to country  $d$  (which corresponds to 100 individuals), leads to a 0.2% increase in content on scientists to Wikipedia. While the coefficients on contributions by registered users are negative and marginally significant, the effect is robust for the *anonymous* contributions to Wikipedia.

As perceived by the Wikipedia community, contributors who skip registration and add content anonymously are less likely to care about their reputation in the community. Anthony et al. (2009) describe two types of anonymous contributors to Wikipedia. The first type are users who see small mistakes in the text and fix them. The second type are experts in their particular fields who come across articles related to their field of expertise and contribute content. As they do not care about their reputation on Wikipedia, they skip the registration procedure and are unlikely to make many contributions to the community. Anthony et al. (2009) show that contributions from anonymous users are as reliable in terms of quality as those of registered users, and have a better quality compared to registered users with few contributions.

Table 6 shows the results for another knowledge relevant domain, “Research institutes” in destination countries. It shows that significantly more content about destination countries is generated by anonymous users in the native languages of immigrants once the inflow of immigrants increases. These results could be interpreted as follows: a 100% increase in the number of migrants from country  $o$  to country  $d$  yields a 5% increase in the number of edits and 27% more content measured in bytes contributed by anonymous users about country  $d$  in the language of country  $o$ . This effect also holds for the other two independent variables of interest, the log immigrants with tertiary education level and an increase in the inflow of immigrants, despite the lower number of observations, and the coefficients are stronger in magnitudes. Similarly to the domain “Scientists”, these results highlight that immigration may attract new occasional Wikipedia users contributing in domains more broadly relevant for technological progress.

To compare these findings to another domain which covers a topic of a broader interest, I show the set of results for the content in the domain “Cuisine” of a destination country (Table 7). The results suggest that similar effects are present. Importantly, along with anonymously contributed content, we can observe growth in content contributed by registered users. The results get stronger in magnitude and significance once skilled immigration is considered. Interestingly, the effects of immigration seem to be distributed more evenly in this domain between anonymous and registered contributions. The fact that the effects hold here as well supports the robustness of the results for knowledge-related domains.

### 3.3 Robustness of baseline results

In this section I perform a set of robustness checks for the baseline results in Tables 8 and 9. Column (1) in Table 8 re-estimates the results for the anonymous edits in column (11) of Table 8 clustering standard errors at the level of origin countries. Similarly, column (1) in Table 9 re-estimates the results in column (1) of Table 9. The magnitude and significance of both coefficients of interest does not change, while standard errors are slightly higher than before.

For both domains, “Scientists” and “Research Institutes”, I further check that the results are not driven by a specific subset of countries. Column (2) shows that the positive effect of interest is robust to the largest destination and origin countries, Germany and the US, and Greece being excluded from the sample. Greece is particularly excluded because of the special significance of Greek ancient science for many cultures, which means that contributions about Greek scientists cannot be attributed to immigration. I also exclude the two largest origin countries, Poland and Romania. The results remain robust. Then, in Column (3) I limit the sample to OECD countries, which have relatively more homogeneous economies and political systems. For the domain “Research Institutes”, without the largest countries, the effect remains at about 5%, while for only OECD countries the magnitude grows to about 8%. For “Scientists”, in column (3) of Table 8 I show that the effect is to a large extent driven by recent years. This is important because we might have expected this large domain to have grown primarily in the early years of Wikipedia. Then, contributions in these early years could be driven by the large amount of information available about the domains being added to Wikipedia by the enthusiasts. However, the fact that the effect is driven by the later years demonstrates that it could be happening due to knowledge dissemination in the process of immigration.

Further, I decompose the regions from which immigrants come into the Middle East, Europe and Asia to study the differences in the effects. The results in Table 8, column (4) show that the positive effect of immigration on anonymously contributed content is driven by immigration from Middle Eastern countries (Azerbaijan, Israel and Turkey), followed by immigration from Asian countries, and is the lowest for Europe (the coefficient approaching to 0.003). For “Research Institutes” the coefficient for European countries is about 0.03.

More important for assessing the robustness of effects is the concern that online content on Wikipedia grows in different languages following patterns of Wikipedia “adoption”, or a learning curve, according to which the language specific community learns how to improve the coverage of topics. In columns (5) - (8) of Tables 8 and 9, I control for potential growth of Wikipedia language editions which is not related to immigration. First, I include linear time trends for each origin and destination country. These trends would account for the growth of content of Wikipedia in the language of country  $o$  about all destination countries, and the growth of content about every destination country  $d$  in all languages. Column (5) for “Scientists” shows that the effects hold, while for “Research Institutes” the significance vanishes as



standard errors grow, potentially, due to inclusion of many variables into the regression with not so many observations, while the coefficient and the standard error make sense. Columns (6) - (8) also control for trends in content generation, including as controls the number of new Wikipedia articles created in each language of Wikipedia (corresponding to countries of origin), total contributions to the domain in the immigrants' native language about all destination countries, and contributions about the domain in English. The robustness checks provide another piece of support to the finding that via immigration more knowledge about destination countries becomes available in the origin country.

Table 8: Immigration Flows and Content Generation on Wikipedia in domain *Scientists*: Robustness checks.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Baseline, SE clust. by origin	Excl. Largest Countries	Period 2014-2016	By Regions	Country time trends	Wikipedia Languages	Native Languages	English Language
$\Delta$ Immigrants, k	0.002*** (0.001)	0.002*** (0.001)	0.011** (0.005)	0.075* (0.042)	0.002** (0.001)	0.002*** (0.001)	0.002*** (0.001)	0.002*** (0.001)
Europe $\times$ $\Delta$ Immigrants, k				-0.072* (0.042)				
Asia $\times$ $\Delta$ Immigrants, k				-0.163*** (0.054)				
Time trend					0.449*** (0.028)	-0.006 (0.022)		
All edits in wikilang.						0.363*** (0.079)		
In native lang., bytes							0.007 (0.026)	
In English lang., bytes								0.057*** (0.021)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Origin country time trends	No	No	No	No	Yes	No	No	No
Host country time trends	No	No	No	No	Yes	Yes	No	No
Mean dep. Variable								
Observations	5394	5394	1269	5394	5394	5379	4653	5394
# of country pairs	698	698	660	698	698	698	603	698
R <sup>2</sup>	0.068	0.068	0.001	0.070	0.153	0.111	0.060	0.070

NOTE: This table contains robustness checks. The independent variable of interest is the log number of immigrants from origin country  $o$  to destination country  $d$  in year  $t$ . All specifications include year dummies. All standard errors (in parentheses) are clustered at the origin country level. Significance stars denote: \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ .

Table 9: Immigration Flows and Content Generation on Wikipedia in domain *Research Institutes*: Robustness checks.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Baseline, SE clust. by origin	Excl. Largest Countries	OECD Countries	By Regions	Country time trends	Wikipedia Languages	Native Languages	English Language
Lg Immigrants	0.048** (0.021)	0.048** (0.021)	0.081*** (0.026)	1.311*** (0.318)	0.036 (0.027)	0.046* (0.024)	0.044** (0.019)	0.049** (0.022)
Europe $\times$ Lg Immigrants				-1.278*** (0.317)				
Asia $\times$ Lg Immigrants				-1.003** (0.391)				
Time trend					-0.017 (0.037)	-0.065* (0.036)		
All edits in wikilang.						0.200* (0.113)		
In native lang., bytes							0.038 (0.051)	
In English lang., bytes								-0.025 (0.034)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Origin country time trends	No	No	No	No	Yes	No	No	No
Host country time trends	No	No	No	No	Yes	Yes	No	No
Mean dep. Variable	0.71	0.71	0.74	0.71	0.71	0.71	0.69	0.71
Observations	822	822	620	822	822	822	754	822
# of country pairs	127	127	94	127	127	127	117	127
R <sup>2</sup>	0.021	0.021	0.026	0.036	0.094	0.062	0.017	0.022

NOTE: This table contains robustness checks. The independent variable of interest is the log number of immigrants from origin country  $o$  to destination country  $d$  in year  $t$ . All specifications include year dummies. All standard errors (in parentheses) are clustered at the origin country level. Significance stars denote: \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ .

### 3.4 Identification challenge: a natural experiment

One potential drawback in the identification of the effect of interest in the baseline analysis is an endogeneity between the measures of immigration and content generated on Wikipedia. The inclusion of country-pair fixed effects addresses this problem by eliminating time-invariant factors that lead to stronger links between countries in terms of both immigration flows and information flows. Also, the inclusion of country-specific trends allows me to control for trends, such as the rising openness of a country, which stimulates higher immigration as well as more information available about this country, including online information on Wikipedia. However, as some residual correlation may remain uncontrolled for, I go further and explore the nature of the immigration phenomenon to obtain better identification of the effect of immigration on knowledge generated online.

Immigration literature increasingly uses global shocks to immigration to identify causal effects of interest (Borjas and Doran (2015), Ganguli (2015), Anelli and Peri (2017), and Barsbai et al. (2017)). For example, Borjas and Doran (2015) and Ganguli (2015) exploit the collapse of Soviet Union in 1991 as a shock in a natural experiment to estimate how the sudden influx of high-skilled immigrants affected scientific publishing. Anelli and Peri (2017) shows that similar shocks to immigration can occur due to macroeconomic conditions. They show that following the economic recession in 2008-2009, emigration from Italy experienced a strong and rather sudden shock in the beginning of 2010. I follow these novel studies and also rely on the fact that economic and political crises trigger emigration from the countries.

I exploit shocks to emigration from origin countries. Within my period of observation, a large subset of origin countries in the sample experienced political and economic crises. Table 13 in the Appendix lists the origin countries that experienced crises, according to the information from Wikipedia articles about each country.\* To show how crisis years in Table 13 correspond to an actual increase in the outflow of migrants from that country, I display total emigration from Russia on Figure 11 in the Appendix. It demonstrates an increase of more than 10,000 persons emigrating from Russia after the political crisis in 2012. Because the observed period in my dataset contains shock to immigration flows, following Angrist and Pischke (2014), the fixed effects estimation (in Tables 5, 6 and 7) could be interpreted causally as difference-in-differences. However, I perform an additional robustness check using a more standard natural experiment setting.

The crisis moment for each immigrant origin country affected by economic or political crisis from Table 13 can be used as a setting for a natural experiment, meaning that a standard difference-in-difference estimation could be performed. Then, I estimate the following equation:

$$Y_{dot} = \alpha_{do} + \alpha_t + \beta_1 T_{do} + \beta_2 T_{do} \times After_t + \beta_3 After_t + \epsilon_{dot}, \quad (2)$$

---

\*According to the rules of Wikipedia all information must have references to its relevant sources, for example, to online media and reports of international organizations.

Table 10 presents the results of equation 2. Positive significant coefficients for countries affected by the crises after the beginning of the crises and the subsequent emigration shock support the evidence from FE estimations: immigration causally increases anonymous contributions to Wikipedia for the domains relevant to research, namely, about scientists and research institutes. After the emigration shock, content in the languages of countries affected by emigration about destination countries grew by about 18% and 24% in terms of edits and 42% and 99% measured in bytes. For the culture-related domain “Cuisine” and research-related domain “Scientists”, contributions both from anonymous and registered users grow, although the magnitude is stronger for anonymously contributed content.

Table 10: Immigration Flows and Content Generation on Wikipedia in all three domains: Difference-in-Differences Estimation.

	Scientists				Research Institutes				Cuisine			
	(1) Anonym. Edits	(2) Reg. Edits	(3) Anonym. Bytes	(4) Reg. Bytes	(5) Anonym. Edits	(6) Reg. Edits	(7) Anonym. Bytes	(8) Reg. Bytes	(9) Anonym. Edits	(10) Reg. Edits	(11) Anonym. Bytes	(12) Reg. Bytes
Treated after T.	0.130** (0.052)	0.062 (0.051)	0.185 (0.144)	0.131 (0.114)	0.236** (0.111)	0.159 (0.135)	0.978** (0.433)	0.248 (0.462)	0.349*** (0.060)	0.293*** (0.058)	0.764*** (0.170)	0.523*** (0.149)
After T.	-0.282*** (0.066)	-0.078 (0.063)	-0.512** (0.211)	-0.104 (0.138)	-0.315* (0.185)	-0.356** (0.157)	-1.105 (0.669)	-1.042** (0.523)	-0.441*** (0.075)	-0.254*** (0.082)	-0.715*** (0.227)	-0.492** (0.209)
Time trend	-0.006 (0.025)	-0.034 (0.024)	0.009 (0.063)	-0.029 (0.060)	-0.036 (0.033)	-0.178* (0.090)	-0.189** (0.095)	-0.416 (0.345)	0.096*** (0.035)	0.040 (0.027)	0.195** (0.082)	0.103 (0.074)
All edits in wikilang.	0.418*** (0.037)	0.566*** (0.049)	1.041*** (0.104)	0.914*** (0.112)	0.154* (0.081)	0.194* (0.104)	0.193 (0.245)	0.309 (0.315)	0.393*** (0.042)	0.380*** (0.040)	0.845*** (0.125)	0.637*** (0.107)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Host country time trends	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean dep. Variable	2.11	3.91	5.19	8.37	0.72	2.04	2.26	5.24	1.83	3.01	4.22	6.46
Observations	5944	5944	5944	5944	818	818	818	818	4828	4828	4828	4828
# of country pairs	689	689	689	689	115	115	115	115	594	594	594	594
R <sup>2</sup>	0.140	0.165	0.088	0.088	0.064	0.102	0.091	0.146	0.117	0.073	0.077	0.056

NOTE: This table contains estimation results for different measures of Wikipedia content (the amount of bytes contributed by anonymous and registered users, the number of edits by anonymous and registered users, minor edits and new articles) about host countries on the languages of origin countries. Columns (1) - (4) show the results for the domain "Scientists", columns (5) - (8) for "Research institutes", and columns (9) - (12) for "Cuisine". All specifications include year dummies. All standard errors (in parentheses) are clustered at the origin-destination country pair level. Significance stars denote: \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ .

## 4 What do anonymous users contribute?

In this section, I further explore, what kind of content contributions were made to Wikipedia by the aforementioned increase in anonymous contributions. I distinguish between edits that are rather minor and edits that add information to the articles. To draw a line between these kinds of edits, I consider that a significant contribution of information is made when one word or more is added. Then, contributions that add more than one word increase the information covered in the article.

As I analyze content generated in a set of languages, I need to consider the length of an average word in each language. I take the maximum average length of one word in a set of languages, 12 characters.\* Then, if the contribution is less than one word, I define it as a minor contribution.

Following the basic specification, I regress the dependent variables, which measure how many minor and substantial edits were anonymously contributed, on the independent variable, the change in the flow of immigrants, which approximates the shock to immigration. In the regressions, I control for origin country time trends in content generation, which capture wiki language-specific trends, as well as for destination country time trends, which stand for topic-specific trends. Table 11 presents the results. We can see that for large domains, “Scientists” and “Cuisine”, small contributions are more strongly affected by shocks to immigration, while for a smaller domain, “Research institutes”, an increase in immigrants yields significant contributions of information to the Wikipedia articles. In terms of the magnitude of these effects, the effect on longer contributions about “Research institutes” in destination countries is the largest: an increase in the immigration flow by a thousand individuals yields a 2% increase in anonymous edits which add at least a full word to Wikipedia.

---

\*Source: <http://www.ravi.io/language-word-lengths>

Table 11: Immigration Flows and the Length of Anonymous Content Generation on Wikipedia in all three domains.

	Scientists		Research Institutes		Cuisine	
	(1) Up to One word	(2) Longer Contribution	(3) Up to One word	(4) Longer Contribution	(5) Up to One word	(6) Longer Contribution
$\Delta$ Immigrants, k	0.003*** (0.001)	0.002*** (0.001)	0.004 (0.012)	0.016* (0.008)	0.005** (0.003)	0.004*** (0.001)
Time trend	0.284*** (0.056)	0.463*** (0.051)	0.016 (0.019)	-0.074** (0.029)	0.171*** (0.060)	0.156*** (0.058)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Origin country time trends	Yes	Yes	Yes	Yes	Yes	Yes
Host country time trends	Yes	Yes	Yes	Yes	Yes	Yes
Mean dep. Variable	1.38	1.58	0.22	0.39	1.32	1.28
Observations	5314	5314	443	443	4277	4277
# of country pairs	668	668	74	74	573	573
R <sup>2</sup>	0.109	0.213	0.089	0.130	0.134	0.152

NOTE: This table contains estimation results for measures of anonymous content added to Wikipedia (minor edits and large edits) about host countries on the languages of origin countries. Columns (1) - (2) show the results for the domain “Scientists”, columns (3) - (4) for “Research institutes”, and columns (5) - (6) for “Cuisine”. All specifications include year dummies, origin country time trends and destination country time trends. All standard errors (in parentheses) are clustered at the origin-destination country pair level. Significance stars denote: \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ .



## 5 Location of anonymous contributions

I further investigate the location of anonymous contributions. For a share of anonymous contributions from each domain I can assign the location based on the IP address of the contributions using R library “rgeolocate”. As this share is not very high, I lose a lot of observations. For example, in the smaller domain “Research institutes” in 2006 only 50% of anonymous contributions are mapped with the countries and cities for IP addresses. This would leave me with about 160 observations in the regressions. Therefore, in what follows I restrict my analysis to the large domains, “Scientists” and “Cuisine”.

Table 12 presents results for the analysis, in which I consider only contributions of immigrants in their native languages about the scientists and cuisine of their destination countries. Further, in Appendix (Table 14) I also show that the results hold if for each destination country we consider contributions about all countries. As the results in Table 12 suggest, as immigration inflow increases, more contributions in the domain “Scientists” and, concretely, about local scientists take place on Wikipedia. This is true for all edits, and the effect is driven by edits which are longer than one word.

Table 12: Immigration Flows and Anonymous Content Generation: Editors Located in Destination Countries and Contributing About Them.

	Scientists				Cuisine			
	(1) Anonym. Edits	(2) Anonym. Bytes	(3) Up to One word	(4) Longer Contribution	(5) Anonym. Edits	(6) Anonym. Bytes	(7) Up to One word	(8) Longer Contribution
$\Delta$ Immigrants, k	0.008** (0.004)	0.002 (0.015)	0.007 (0.005)	0.008** (0.004)	-0.007 (0.008)	-0.008 (0.029)	-0.007 (0.005)	-0.006 (0.008)
Time trend	0.212** (0.093)	0.332 (0.294)	0.211* (0.119)	0.127** (0.055)	-0.085 (0.088)	0.579 (0.370)	-0.103** (0.041)	0.015 (0.110)
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Host country time trends	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Origin country time trends	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean dep. Variable	2.06	5.27	0.97	1.88	1.48	4.22	0.76	1.16
Observations	881	881	881	881	791	791	791	791
# of country pairs	180	180	180	180	152	152	152	152
R <sup>2</sup>	0.242	0.209	0.185	0.236	0.135	0.144	0.104	0.134

NOTE: This table contains estimation results for different measures of Wikipedia content (the amount of bytes contributed by anonymous and registered users, the number of edits by anonymous and registered users, minor edits and new articles) about host countries on the languages of origin countries. Columns (1) - (6) show the results where the independent variable of interest is log number of immigrants from origin country  $o$  to destination country  $d$  in year  $t$ . In columns (7) - (8) the independent variable of interest is the log number of immigrants with tertiary education. All specifications include year dummies. All standard errors (in parentheses) are clustered at the origin country level. Significance stars denote: \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ .

## 6 Concluding remarks

This paper sheds light on the impact of immigration flows on the dissemination of broader knowledge via one of the world’s most viewed online knowledge repositories, Wikipedia. It uses shocks to immigration to identify the causal effect of immigration inflows on contributions to Wikipedia about host countries in the languages of immigrants’ origin countries. The nature of Wikipedia makes it possible to investigate its content development using numerous content measures. I specifically compute the total edits and bytes (symbols) contributed by registered users as well as anonymous contributors who skipped the log-in procedure before contributing. For the anonymous contributions, I mapped the IP-addresses of contributions with the destination countries of immigrants to enhance the precision of the results.

My findings suggest that more online knowledge becomes available about the host countries in the languages of origin countries on Wikipedia when shocks to immigration occur. Contributions to the domains relevant for science, “Scientists” and “Research institutes”, are driven by an increase in the contributing activity of anonymous (unregistered) users, who, in line with Wikipedia philosophy, may be either occasional contributors or experts in their fields. As studies of the Wikipedia community indicate anonymous users are likely to occasionally view Wikipedia articles in their area of interest and contribute just once or a few times. Moreover, when we consider the location of anonymous edits, the results become more precise and suggest that the increase in anonymous content is driven by longer anonymous contributions which add more than a word to Wikipedia articles.

Knowledge disseminated due to immigration processes might have numerous implications for research, external research contacts, technology adoption and individual knowledge-related choices and can therefore shape the future development of immigrants’ origin countries. This study highlights the importance of information and communication technologies in facilitating cross-border knowledge exchange and mitigation of the negative effects of brain drain in less developed countries.

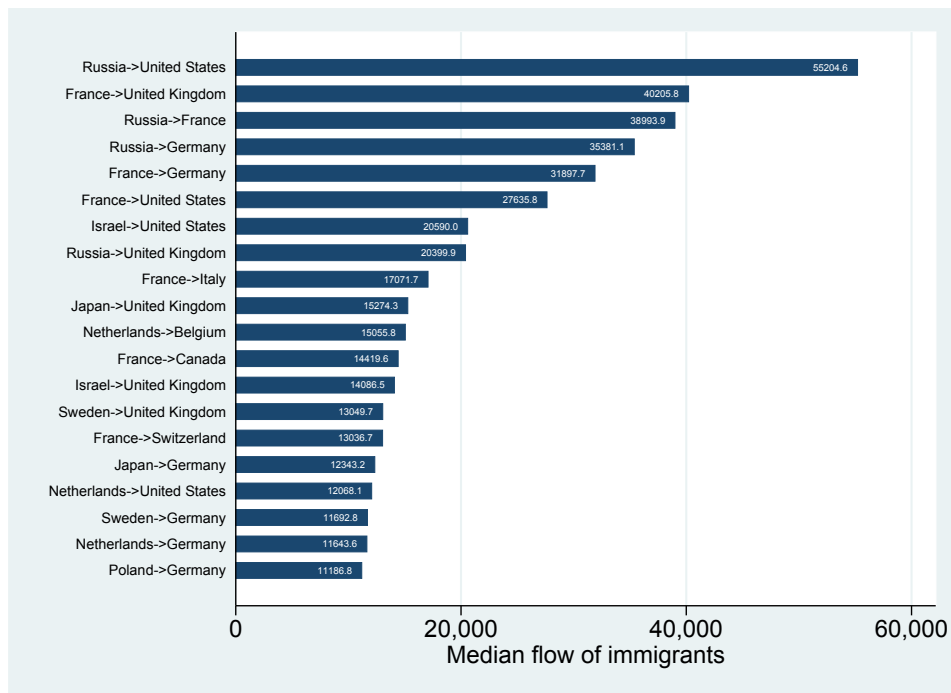
## References

- Agrawal, Ajay, Devesh Kapur, John McHale, and Alexander Oettl**, “Brain drain or brain bank? The impact of skilled emigration on poor-country innovation,” *Journal of Urban Economics*, 2011, 69 (1), 43–55.
- , **Nicola Lacetera, and Elizabeth Lyons**, “How Do Online Platforms Flatten Markets for Contract Labor?,” Technical Report, Working Paper 2012.
- Anelli, Massimo and Giovanni Peri**, “Does emigration delay political change? Evidence from Italy during the great recession,” *Economic Policy*, 2017, 32 (91), 551–596.
- Angrist, Joshua D and Jörn-Steffen Pischke**, *Mastering’metrics: The path from cause to effect*, Princeton University Press, 2014.
- Anthony, Denise, Sean W Smith, and Timothy Williamson**, “Reputation and reliability in collective goods: The case of the online encyclopedia Wikipedia,” *Rationality and Society*, 2009, 21 (3), 283–306.
- Asatryan, Zareh, Benjamin Bittschi, and Philipp Doerrenberg**, “Remittances and public finances: Evidence from oil-price shocks,” *Journal of Public Economics*, 2017.
- Barsbai, Toman, Hillel Rapoport, Andreas Steinmayr, and Christoph Trebesch**, “The effect of labor migration on the diffusion of democracy: evidence from a former Soviet Republic,” *American Economic Journal: Applied Economics*, 2017, 9 (3), 36–69.
- Borjas, George J and Kirk B Doran**, “Cognitive mobility: Labor market responses to supply shocks in the space of ideas,” *Journal of Labor Economics*, 2015, 33 (S1), S109–S145.
- Bosetti, Valentina, Cristina Cattaneo, and Elena Verdolini**, “Migration of skilled workers and innovation: A European Perspective,” *Journal of International Economics*, 2015, 96 (2), 311–322.
- Breschi, Stefano and Francesco Lissoni**, “Mobility of skilled workers and co-invention networks: an anatomy of localized knowledge flows,” *Journal of Economic Geography*, 2009, 9 (4), 439–468.
- Docquier, Frédéric, Elisabetta Lodigiani, Hillel Rapoport, and Maurice Schiff**, “Emigration and democracy,” *Journal of Development Economics*, 2016, 120, 209–223.
- Douglas, Kacey N**, “International knowledge flows and technological advance: the role of migration,” *IZA Journal of Migration*, 2015, 4 (1), 1.
- Fackler, Thomas, Yvonne Giesing, and Nadzeya Laurentsyeva**, “Knowledge Remittances: How Emigration Fosters Innovation in Source Countries,” 2016.
- Ganguli, Ina**, “Immigration and Ideas: What Did Russian Scientists Bring to the United States?,” *Journal of Labor Economics*, 2015, 33 (S1 Part 2), S257–S288.
- Ghani, Ejaz, William R Kerr, and Christopher Stanton**, “Diasporas and outsourcing: evidence from oDesk and India,” *Management Science*, 2014, 60 (7), 1677–1697.
- Gould, David M**, “Immigrant links to the home country: empirical implications for US bilateral trade flows,” *The Review of Economics and Statistics*, 1994, pp. 302–316.
- Hergueux, Jérôme, Yann Algan, Yochai Benkler, and Mayo Fuster Morell**, “Cooperation in Peer Production Economy: Experimental Evidence from Wikipedia,” in “Lyon Meeting” 2014.
- Hinnosaar, Marit, Toomas Hinnosaar, Michael E Kummer, and Olga Slivko**, “Wikipedia Matters,” 2017.
- Hunt, Jennifer and Marjolaine Gauthier-Loiselle**, “How much does immigration boost innovation?,” *American Economic Journal: Macroeconomics*, 2010, 2 (2), 31–56.
- Javorcik, Beata S, Çağlar Özden, Mariana Spatareanu, and Cristina Neagu**, “Migrant networks and foreign direct investment,” *Journal of Development Economics*, 2011, 94 (2), 231–241.
- Kerr, William R**, “Ethnic scientific communities and international technology diffusion,” *The Review of Economics and Statistics*, 2008, 90 (3), 518–537.
- Kugler, Maurice and Hillel Rapoport**, “International labor and capital flows: Complements or substitutes?,” *Economics Letters*, 2007, 94 (2), 155–162.
- Kummer, Michael E**, “Spillovers in Networks of User Generated Content,” *Available at SSRN*, 2013.
- Migueluez, Ernest**, “Inventor diasporas and the internationalization of technology,” *The World Bank Economic Review*, 2016, p. lhw013.

- **and Claudia Noumedem Temgoua**, “Highly Skilled Migration and Knowledge Diffusion: A Gravity Model Approach.”
- Peri, Giovanni**, “Determinants of knowledge flows and their effect on innovation,” *Review of Economics and Statistics*, 2005, 87 (2), 308–322.
- Piskorski, Mikolaj Jan and Andreea D Gorbatai**, “Testing Colemans social-norm enforcement mechanism: Evidence from Wikipedia,” *Harvard Business School Strategy Unit Working Paper*, 2013, (11-055).
- Slivko, Olga**, “Peer effects in collaborative content generation: The evidence from German Wikipedia,” *ZEW-Centre for European Economic Research Discussion Paper*, 2014, (14-128).
- Thompson, Neil and Douglas Hanley**, “Science Is Shaped by Wikipedia: Evidence from a Randomized Control Trial,” 2017.
- Zhang, Xiaoquan Michael and Feng Zhu**, “Group size and incentives to contribute: A natural experiment at Chinese Wikipedia,” *American Economic Review*, 2011, 101 (4), 1601–1615.

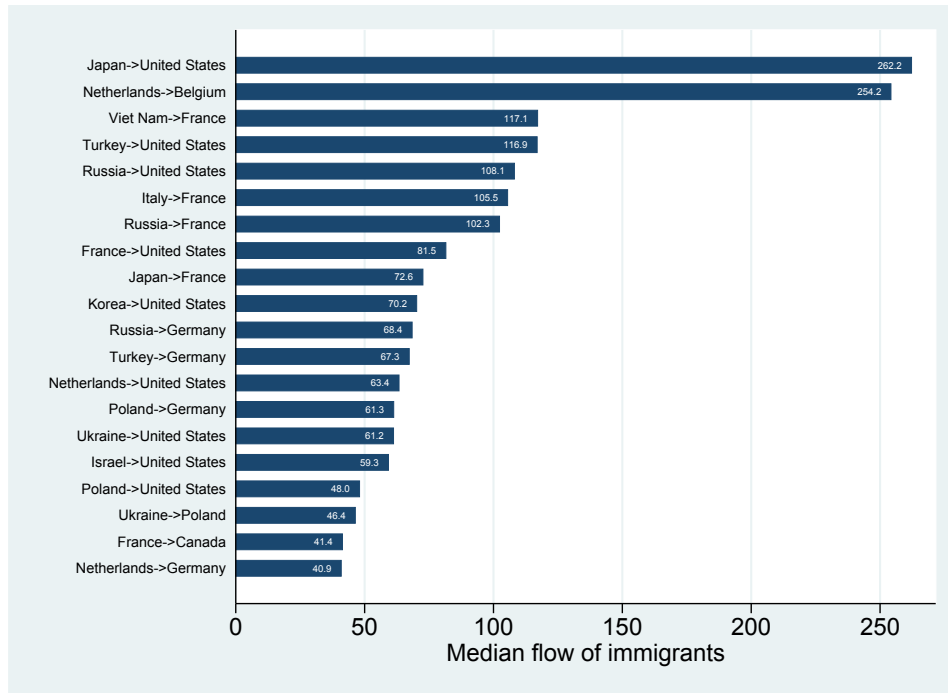
## A Appendix

Figure 6: Country pairs where immigrants from origin countries contribute most knowledge about destination countries over 2006-2015: domain "Scientists".



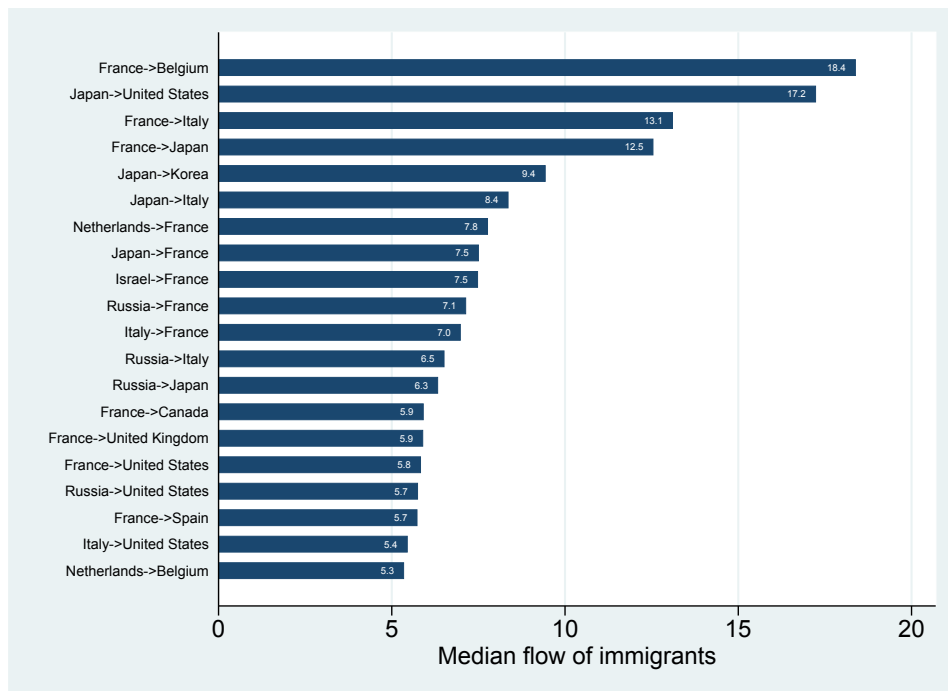
NOTE: This figure illustrates top 20 Wikipedia content flows in the domain "Scientists", where knowledge is contributed about destination country in the language or immigrant origin country. The bar labels display the pair of origin -> destination countries.

Figure 7: Country pairs where immigrants from origin countries contribute most knowledge about destination countries over 2006-2015: domain "Research Institutions".



NOTE: This figure illustrates top 20 Wikipedia content flows in the domain "Research Institutions", where knowledge is contributed about destination country in the language or immigrant origin country. The bar labels display the pair of origin -> destination countries.

Figure 8: Country pairs where immigrants from origin countries contribute most knowledge about destination countries over 2006-2015: domain "Cuisine".



NOTE: This figure illustrates top 20 Wikipedia content flows in the domain "Cuisine", where knowledge is contributed about destination country in the language or immigrant origin country. The bar labels display the pair of origin -> destination countries.

Table 13: The years of crises for origin countries.

	Year of Crisis
Albania	2012
Armenia	x
Azerbaijan	x
Belarus	2011
Bulgaria	2009
Croatia	2011
Czech Republic	2009
Denmark	x
Estonia	x
Finland	x
France	x
Georgia	2008
Greece	2009
Hungary	2008
Iceland	2008
Israel	x
Italy	2009
Japan	x
Kazakhstan	x
Korea	x
Latvia	2008
Lithuania	x
Macedonia	x
Mongolia	2009
Netherlands	x
Norway	x
Poland	x
Romania	2008
Russia	2012
Serbia	2009
Slovak Republic	x
Slovenia	2009
Sweden	x
Thailand	x
Turkey	x
Ukraine	2008
Viet Nam	x



Figure 9: Example of immigration from Spain and Netherlands and content contributions to Wikipedia about German Max Planck Society

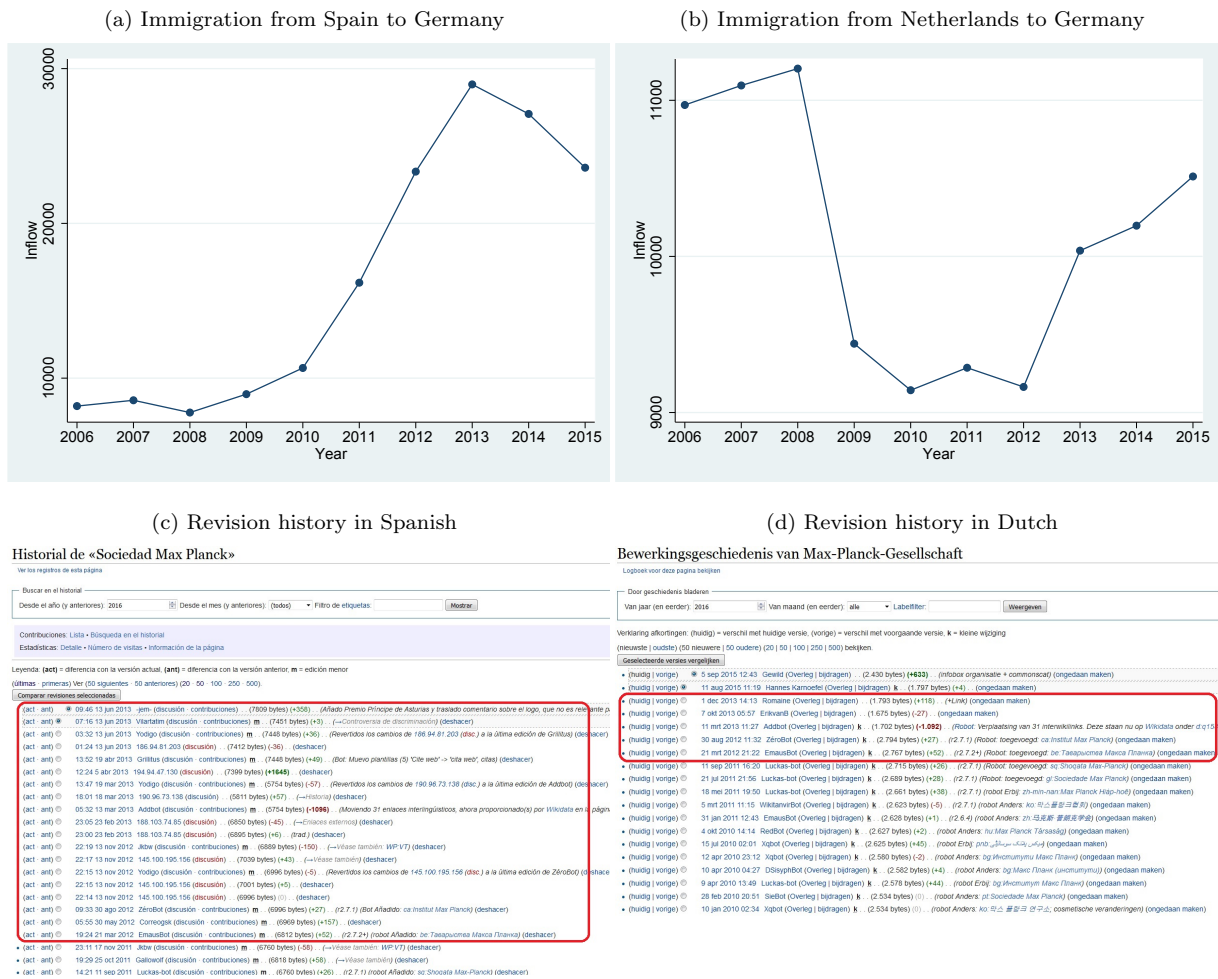
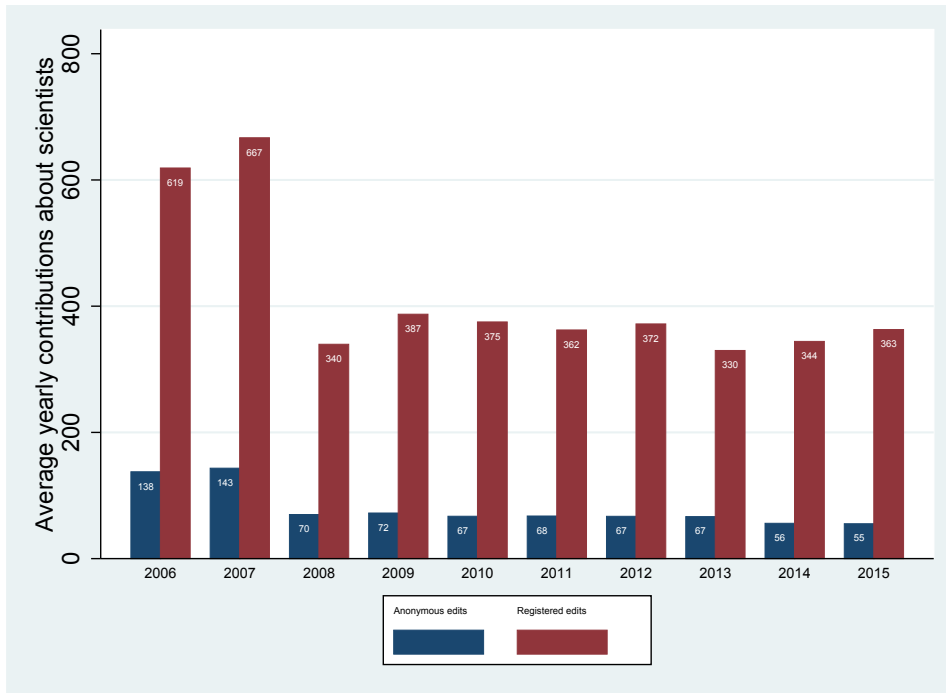
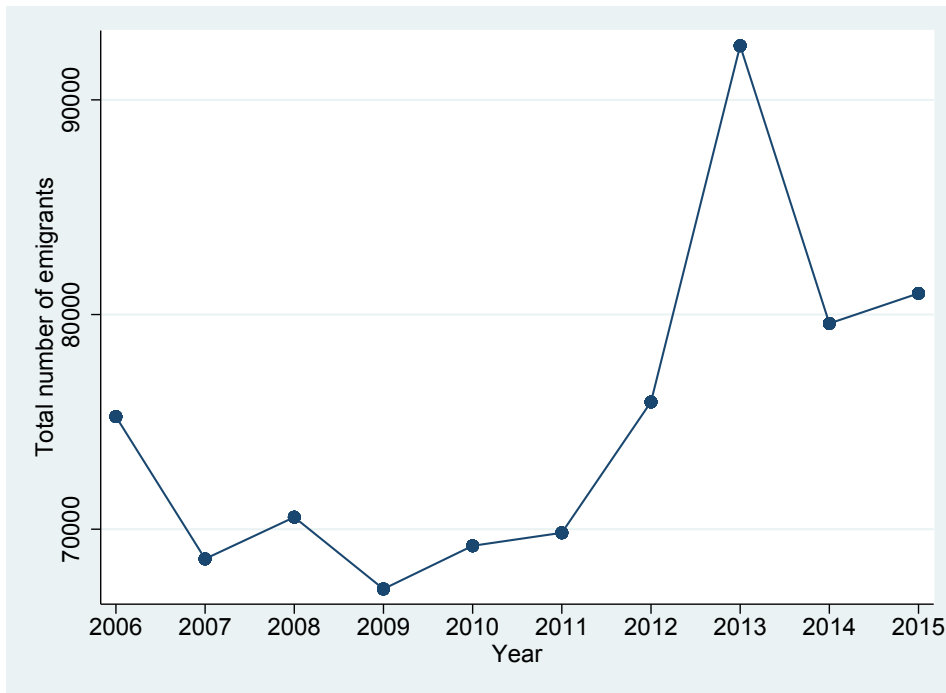


Figure 10: Average contributions to the domain “Scientists” over years.



NOTE: This figure displays average yearly edits about “Scientists” by immigrants coming from an origin country to a destination country. Each pair of bars shows the yearly mean edits performed by anonymous and registered contributors on Wikipedia.

Figure 11: Emigration from Russia after the political crisis in 2012.



NOTE: This figure illustrates how total emigration from Russia to OECD countries changes after the political crisis in 2012.

Table 14: Immigration Flows and Anonymous Content Generation: Editors Locate in Destination Countries and Contribute About All Countries.

	Scientists				Cuisine			
	(1) Anonym. Edits	(2) Anonym. Bytes	(3) Up to One word	(4) Longer Contribution	(5) Anonym. Edits	(6) Anonym. Bytes	(7) Up to One word	(8) Longer Contribution
$\Delta$ Immigrants, k	0.002** (0.001)	-0.004 (0.005)	-0.001 (0.002)	0.003*** (0.001)	0.003 (0.005)	0.006 (0.013)	0.004 (0.005)	0.003 (0.004)
Time trend	0.057 (0.044)	0.529** (0.244)	0.037 (0.040)	0.044 (0.036)	0.033 (0.041)	0.520*** (0.123)	0.015 (0.100)	0.109* (0.060)
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Host country time trends	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Origin country time trends	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean dep. Variable	1.90	5.20	0.85	1.71	1.85	5.06	1.08	1.49
Observations	3286	3286	3286	3286	1928	1928	1928	1928
# of country pairs	597	597	597	597	331	331	331	331
R <sup>2</sup>	0.173	0.135	0.083	0.188	0.119	0.118	0.081	0.140

NOTE: This table contains estimation results for different measures of Wikipedia content (the amount of bytes contributed by anonymous and registered users, the number of edits by anonymous and registered users, minor edits and new articles) about host countries on the languages of origin countries. Columns (1) - (6) show the results where the independent variable of interest is log number of immigrants from origin country  $o$  to destination country  $d$  in year  $t$ . In columns (7) - (8) the independent variable of interest is the log number of immigrants with tertiary education. All specifications include year dummies. All standard errors (in parentheses) are clustered at the origin country level. Significance stars denote: \*\*\*  $p < 0.01$  , \*\*  $p < 0.05$  , \*  $p < 0.1$  .