Do Immigrants Return Knowledge Home? The Evidence on Knowledge Dissemination via Wikipedia*

Olga Slivko**

January 10, 2017

Abstract

While previous studies highlight the positive impact of immigration on cross-border patenting and scientific publications, the role of immigration flows in the dissemination of knowledge in a wider context is not fully assessed. In this paper, I estimate the effect of immigration on the facilitation of online knowledge in relevant domains. To quantify online knowledge, I focus on one of the world's most viewed knowledge platforms, Wikipedia.

I combine the data on (skilled) immigration flows between the pairs of countries of immigrants' origin and destination with contributions to Wikipedia about destination countries in the native languages of origin countries. The knowledge domains I look at are related to science, technology and culture. In order to draw a causal inference, I use spikes in immigration above 30% in the origin countries as exogenous shocks with respect to Wikipedia content and analyze subsequent changes in the rates of contribution to Wikipedia in a difference-in-differences framework. The results suggest that an increase in immigration yields more knowledge contributed to Wikipedia about destination countries on the native languages of the origin countries. The increase in contributions stems mostly from anonymous (potentially, new or occasional) contributors. Moreover, once spikes in skilled immigration are considered, the effects on science and technology domains get stronger.

Keywords: Skilled Immigration; Online knowledge; Wikipedia.

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

* I benefitted very much from the advice of Michael Ward, Patrick Schulte, Irene Bertschek, Mary O'Mahony, Michael Kummer, Marianne Saam, Michael Zhang, Vivek Ghosal, Daniel Erdsiek, Reinhold Kesler, Thomas Niebel, Fabiene Rasel, Steffen Viete, Stefano Castriota. I am also indebted to Tobias Werner, Niklas Isaak and Lukas Trotner for their excellent research assistance. I thank 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*

1 Introduction

Immigration fosters competitiveness and technological progress in the host countries by contributing to more dynamic knowledge creation and dissemination. As the recent article in The Independent pointed, "All six of America's 2016 Nobel Prize winners are immigrants"*. Economic studies suggest that indeed, a substantial share of the innovative potential of countries comes from the outside (Peri (2005)), and immigration of skilled employees and scientists is an important channel for that. Immigration is positively related to product innovation (Ozgen et al. (2014)) as well as with patenting and scientific knowledge generation (Hunt and Gauthier-Loiselle (2010), Bosetti et al. (2015), and Ganguli (2015)).

In contrast, for the countries of immigrants' origin the outflow of skilled individuals is often perceived as an eroding phenomenon. However, recent literature on innovation suggests that knowledge spills over beyond frontiers and immigrants can act as a channel through which their origin countries could benefit from the external knowledge to a certain extent. One of the reasons for that channel could be that, as shown in Grogger and Hanson (2011), Chiquiar and Hanson (2002), and Foley and Kerr (2013), there is a positive selection of migrants to the destination countries with regard to education and skills, which means that individuals coming to new destinations might be more active in knowledge acquisition (and dissemination) than their non-moving counterparts in the country of origin and the country of destination. In this study, I address this potential channel of knowledge dissemination focusing on Wikipedia, the world's largest online encyclopedia. For that, I study the impact of immigration on the availability of meta knowledge about the host countries in the languages of immigrants' origin countries.

The positive impact of immigration on knowledge spillovers to the countries of immigrants' origin have been shown for scientific publications and inventions. Evidence sugests that, as immigration takes place to the US, more patents from the US get cited by and cite the patents in countries from where immigrants come (Douglas (2015), Ganguli (2015)). In Fackler et al. (2016) knowledge spillovers flow in both directions: the inflow of high-skilled immigrants to the destination countries enhances patent applications in host countries, which cite prior art patents from the origin countries, and vice versa. Since these results only focus on the spillovers to inventions and scientific publications, they might capture the lower bound of the overall knowledge exchange taking place due to immigration. In fact, a wider availability of information and communication technologies, such as various online cites and access to the Internet using laptops and smartphones, helps individuals to become facilitators of knowledge. Moreover, immigrants proceeding from their origin countries could be able to find a better form to express knowledge such that it could be easier absorbed in their former fatherland. These factors suggest that the amount of general knowledge disseminated due to immigration could be far beyond formal scientific knowledge.

For studying meta knowledge exchange between countries due to immigration I combine data on

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

immigration flows between some European and Asian countries available from OECD and contributions to Wikipedia in the specific knowledge domains in a set of languages. I exploit the fact that some categories on Wikipedia cover country-specific topics, for instance, "Research Institutes" in a country, or "Cuisine" of a country in multiple languages, including languages of immigrant origin countries from my sample. Then, to identify the effects of interest, I use the spikes in immigration (increases in immigration flows of more than 30% from a year to another) that stem from shocks arguably exogenous to Wikipedia content, potentially due to economic and political crises. Comparing treated immigration pairs with control pairs in a difference-in-differences approach, I analyze the respective changes in contributions to content about destination countries on the languages of origin countries before and after the shocks to immigration took place.

The results suggest that, after immigration rises, more edits by unregistered contributors take place on Wikipedia about the research institutions, software and cuisine in destination countries on the languages of origin countries on Wikipedia. In the year of the shock, a 1 per cent increase in immigration yields a 0.28-0.62 per cent increase in edits performed by anonymous or registered contributors. This is an institutional feature of Wikipedia that individuals can optionally edit articles skipping the registration and authorization procedure or after logging in. With immigration, anonymous edits grow in the domains of research institutes and cuisine, while registered edits grow in the domain on software. While shocks to overall immigration bring more editing to topics on cuisine and software, limiting considered spikes in immigration only to the increases in high skilled immigrant flows yields stronger effects to edits about research institutes, stronger but less significant effects to edits about software and insignificant effects to edits about cuisine. Surprisingly, while immigration has a positive effect on editing activity, the amount of bytes of information contributed does not increase, which might suggest an improvement of quality of articles rather than the quantity of information on Wikipedia.

These findings add to the literature on incentives for contribution to public goods, including online public goods and, in particular, Wikipedia. Previous 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 or themselves reinforce social spillovers on Wikipedia.

Information contributed to Wikipedia can, to some extent, have implications for the technology adoption and individual knowledge-related choices. I find that, in particular, more information available about science and technology of the host countries is positively related to the future international students' choices for the countries of destination to spend their semesters abroad. A better understanding of the impact that immigration could have on knowledge dissemination to the source countries has potentially

important implications for policy makers. On the contrary to a widely acknowledged hampering effect of outflow of human capital from the origin country, more knowledge about external technological opportunities in the origin countries could contribute to promoting the adoption of more advanced technologies and new business opportunities. In this aspect, my study relates to Asatryan et al. (2016) who point that immigrants contribute to their origin countries by money remittances to their families and, thus, positively affecting consumption in origin countries.

The remainder of the paper is structured as follows. Section 2 describes the data set. Section 3 discusses the empirical approach and reports the results on the direct effect of immigration on knowledge dissemination. Section 4 investigates whether the opposite effect takes place, such that immigrants disseminate knowledge about their origin countries in the host countries. Section 5 provides some robustness checks. Section 6 investigates potential implications of knowledge dissemination via Wikipedia on student mobility choices. Finally, Section 7 concludes.

2 Data and Descriptive Statistics

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 includes for source-destination country pairs indicators for inflow, outflow and stock of migrants, covering years 2000 - 2014. I select a subset of countries, including European and Asian countries, amounting to 21 origin countries and 12 host country. I exclude countries with potentially large illegal immigration (some Middle East countries) and those with the languages spoken in many other countries, such as English, Spanish, Persian or Arabic. As the immigrant data are further merged with the data on online content generation from Wikipedia in years 2006-2015, I use the immigration inflow data over the years 2006-2014.

The OECD International Migration Database provides yearly data, but without further distinguishing education or skill levels of migrants. Therefore, I extend it with data on brain drain from IAB* where for 20 OECD countries the total number of foreign-born individuals aged 25 years and older, living in each of the 20 considered OECD destination countries, is decomposed by skills into high (tertiary), medium (high-school leaving certificate) and low (lower secondary, primary and no schooling). Based on these numbers I compute shares of high-, medium- and low-skilled migrants to apply them to OECD immigration inflow data. However, the data with skills decomposition are only available for the years 2005 and 2010. Therefore, for each year between 2005 and 2010 I impute the average share between the two bounding available years, 2010 is present in the data, and for the years after 2010 I apply 2010's shares as the data of 2015 are not available. Further, I multiply these shares with total inflows from OECD data to assess the approximate flows of immigrants by the level of education.

*Institute für Arbeitsmarkt- and Berufsforschung, wwww.iab.de, for further detail see Bruecker et al. (2013)

Further, I merge data on immigration with the measure of knowledge dissemination. As Wikipedia is one of the world's most viewed online knowledge repositories, I use it as such a measure for online knowledge. For combining relevant knowledge on Wikipedia with immigration flows, I exploit a specific feature of Wikipedia, namely, the article categories, which encompass all articles on Wikipedia belonging to the same knowledge domains. Among these categories, there are country-specific categories of articles that focus on some large knowledge topics, for example, "Research Institutes" or "Cuisine" of each country. Hence, these country-specific categories can represent the available knowledge about the host countries. As an example, "Research Institutes in Germany" includes articles about German research centers and researchers working in Germany. Moreover, articles, belonging to these categories exist in many languages. This allows to relate knowledge available about host countries with the language of the origin country with the immigration data for origin-destination country pairs.

As measures of Wikipedia content, I use the number of edits to articles, total and distinguishing whether the contributor logged in or made the edit anonymously, and the number of contributors who appear as registered users for the first time in my sample, i.e. new registered users, aggregating all indicators over years. The data from Wikipedia are collected via API-box tool for the years 2006-2015.

To assess the sample resulting from merging data on immigration flows with online knowledge contributions to Wikipedia, Table (1) provides data aggregated at the level of host countries. In columns, it displays total immigrant inflows together with the size of added content and newly created pages in two knowledge domains, research institutes and cuisine. All added content refers to the host countries and is aggregated across all languages of source countries in the sample. We can observe in the table large immigration inflows for countries like Germany, Italy or Japan are associated with large amounts of content added about the cuisine of the country as well as about science and technology country-specific topics.

To shed some light on how knowledge on Wikipedia might matter of knowledge economies of the origin countries, in the last Section I analyze the relation between the availability of knowledge about science and technology on student mobility choices. For that, I merge Wikipedia content indicators with the data on inbound international student mobility from Unesco Institute of Statistics *, which provide number of students moving for studies from origin to host countries. Table (2) displays descriptive statistics for the variables used as dependent and explanatory in the remaining analyses. Immigration and Wikipedia content measures appear to be highly skewed so I take their natural logarithms for the further quantitative analysis.

*Source: http://data.uis.unesco.org/

Table 1: Total immigration inflows and the number of anonymous and registered edits about host countries in categories and over the years 2006-2014.

	# Immigrants (K)	Sci&Tech., Anonym.	Sci&Tech., Reg.	Cuisine, Anonym.	Cuisine, Reg.
Chile	47.05	36291	54325	21874	28821
Czech Republic	300.93	3811	16864	2556	5978
Denmark	181.59	14941	54448	2100	5788
Estonia	2.93	4633	15900	436	1260
Finland	94.79	3771	12639	1909	4778
France	186.88	33636	146125	18668	62466
Germany	4405.21	158064	414228	12015	29425
Greece	234.00	11350	47969	5583	15723
Hungary	151.60	31815	85116	3566	9251
Iceland	32.07	986	5028	394	1476
Ireland	0.00	195	1263	978	3788
Israel	20.44	6542	25771	4076	6288
Italy	1730.79	86807	240018	34308	84134
Japan	1502.58	30164	82505	19908	59661
Korea	1623.36	1651	5733	8765	21844
Latvia	7.92	226	1477	172	538
Netherlands	489.86	40725	108074	2803	8572
Norway	326.71	9855	42655	1454	4005
Poland	244.14	19547	76584	5538	12548
Portugal	88.40	1133	4676	1468	3221
Slovak Republic	68.28	1874	8563	1248	3333
Slovenia	81.23	2295	6989	409	814
Spain	1370.90	10003	34390	7165	21320
Sweden	313.56	61448	209482	2847	7570
Turkey	13.57	3641	16628	10196	20791
Total	13518.81	575404	1717450	170436	423393

For each destination country columns (1)-(5) display the total number of immigrants as well as anonymous and registered edits about destination countries aggregated across all countries of immigrants' origin.

Table 2: Summary statistics: all variables.

	Mean	Std. dev.	Minimum	Maximum
Immigrants	3488	15238	0	271443
Share of high-skilled immigrants	.34	.14	.043	.75
High-skill. immigr.	1077	3773	0	53252
Total articles in language, k	413	393	3.3	1700
Research institutes, edits	79	162	0	1014
Research institutes, an.edits	20	48	0	334
Research institutes, reg.edits	59	116	0	685
Research institutes, new editors	21	49	0	366
Software, edits	334	1745	0	15664
Software, an.edits	170	975	0	9715
Software, reg.edits	165	779	0	7108
Software, new editors	88	394	0	3431
Cuisine, edits	174	442	0	6161
Cuisine, an.edits	49	163	0	2917
Cuisine, reg.edits	125	294	0	4857
Cuisine, new editors	45	128	0	1938

3 Do immigrants disseminate knowledge to their home countries?

3.1 Empirical strategy

Immigration flows can act as a powerful channel for knowledge dissemination. In this study, I analyze the relationship between immigration and knowledge dissemination on Wikipedia in two frameworks. First, I use panel data OLS to establish the relationship between inflow of high-skilled immigrants from a number of origin countries and knowledge generated about host countries on the languages of those origin countries. Second, in the main specification, I address the potential endogeneity between bilateral immigration flows and the amount of online content available about one country on the language of another one. I use the fact that between some country pairs immigration flows increase in one year by more than 30%, which could be caused, for instance, by political or economic crises.

3.1.1 OLS Regression

Before conducting full difference-in-differences analysis I estimate the panel data OLS specification to test whether there is a positive relationship between immigration flows and content contributions to

Wikipedia. For that, I regress the amount of content generated on Wikipedia about host countries on the languages of origin countries on the number of high-skilled immigrants. The regression equation is given by:

$$Content_{dot} = \alpha\ ImmigrationInflow_{dot} + \beta\ Controls_{ot} + \gamma_{do} + \delta_t + \epsilon_{dot}, \tag{1}$$

where d stands for the country of destination or the topic dedicated to the host country on Wikipedia, o for the country of immigrants' origin or, for Wikipedia data, the language of the content, t is the current year. Country pair fixed effects are included to control for the time-invariant heterogeneity, for example, for the popular migration destinations for every origin country and so the online content availability, for example, for the content about Germany in Turkish language. Time (year) effects allow to control for common time trends in online content generation. To control for the fact that some language editions of Wikipedia grow faster than the others, and this could affect the increase in the amount of information on some particular category, I add as a control variable the number of articles available in the language of country of origin o in year t. Its dynamics captures the development in time of every language edition of Wikipedia due to other reasons than immigration (for instance, spillovers on the platform). In all estimations, I cluster standard errors by origin-destination country pairs to allow for serial correlations in the bilateral immigration flows.

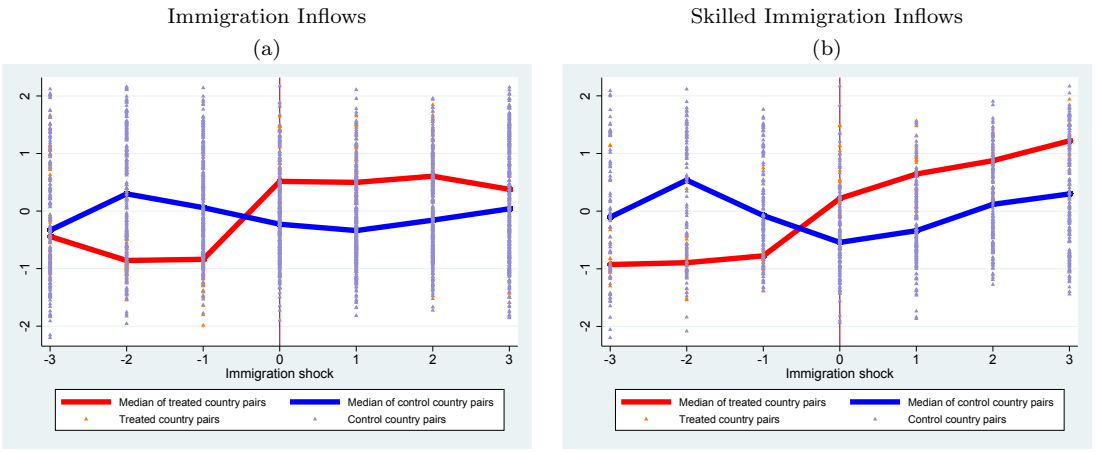
3.1.2 Difference-in-Differences

As an alternative to OLS, I exploit the causal relationship between immigration and knowledge dissemination in the difference-in-differences (DiD) approach. I use the fact that immigration flows between some country pairs increase by more than 30% over the year, which comes as sudden with respect to the available online content and could be caused, for instance, by political or economic crises. To define affected country pairs I use a threshold of 30% for an increase in immigration flow over one year, and that the median immigration flow between the countries should be sufficiently large (above the sample median) such that a large increase in the bilateral flow could yield a significant change in the number of immigrants from the origin country in the host country.

The resulting groups of bilateral immigration flows defined based on this criterion are depicted in Figure 1. It shows the difference in median immigration flows between the groups of treated and control country pairs over time. The moment of an increase in immigration inflow is in point 0 of the time line. The left hand shows treated and control bilateral inflows when the spikes in immigration are defined based on an increase in the total immigration flows. In contrast, on the right hand the shock is defined based on spikes in high-skilled immigration. All the values of immigration flows are normalized with respect to the mean and standard deviation of each country pair.

The differences in online content generation rates between origin-destination country pairs affected

Figure 1: Comparison of immigration outflows between the groups of treated and control country pairs over time with respect to the moment of the shocks.



NOTE: This figure shows the difference between immigration flows between the groups of country pairs treated by the spike in bilateral immigration and controls before and after the spike. On the left hand figure (a) the spikes in immigration are defined as an increase in the immigration flow higher than 30%. In contrast, on the right hand figure (b) the spikes in immigration are defined based on high-skilled immigration. The values of immigration flows are normalized with respect to the mean and standard deviation.

by the crisis and controls are displayed on Figure 2. In rows, the results are presented by knowledge domains and in columns by the measure of content added to Wikipedia, the number of edits performed by anonymous and registered contributors. All graphs demonstrate moderate differences between the group of affected and unaffected country pairs after the shocks take place. To verify whether these differences are statistically significant controlling for individual heterogeneity, I further conduct difference-in-difference estimation.

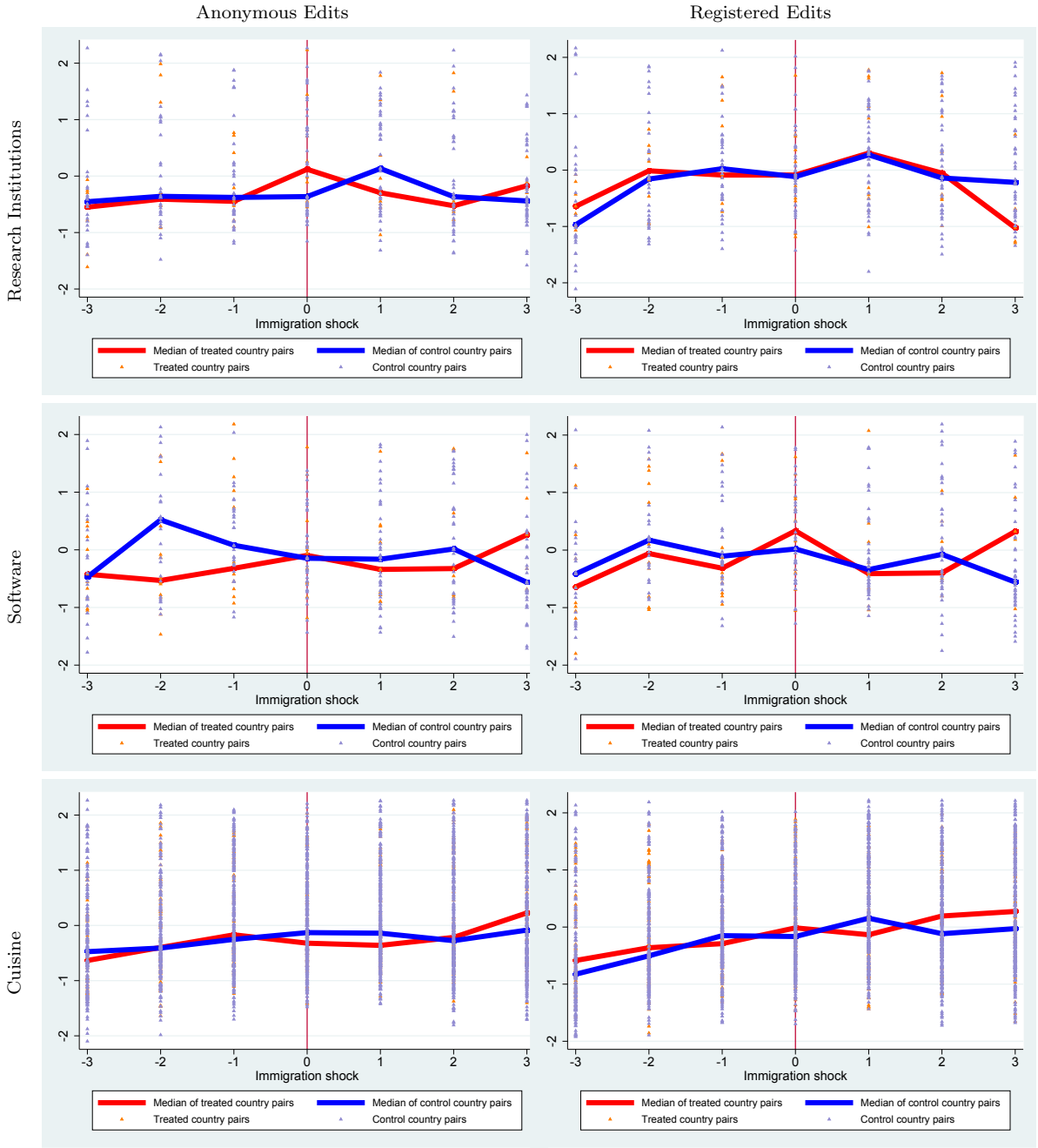
The DiD approach, in the the first difference, compares the amount of content generated about host countries on the languages of origin countries before and after the shocks, while in the second difference it compares content generation in treated country pairs to control country pairs, for which immigration inflows did not experience spikes. The assumption for using DiD approach is that there are no time-variant factors which would affect the spikes to immigration between countries as well as content generated on Wikipedia. All factors which could be considered time-invariant over the observed period, such as, for example, the strength of country-to-country cultural links, are controlled with this approach.

The difference-in-difference regression is:

$$WikiContent_{dot} = \alpha \ (Treated_{do} \times \sum D(Year = t)) + \beta \ Controls_{ot} + \gamma_{do} + \delta_t + \epsilon_{dot}, \qquad (2)$$

$Treated_{do}$ is a dummy variable for bilateral immigration flows treated by the spikes. Since it does not vary over time, it drops out in the fixed-effects specification. The coefficient of interest for the treatment effect is α, which stands for the interaction terms of treated pairs and year dummies. This specification allows to decompose our treatment effect of interest by years subsequent to the spike.

Figure 2: Comparison of Wikipedia content generation for affected and unaffected country pairs over time with respect to the moment of the shocks.



Note: This figure displays the median trends in online content generated before and after the shocks to immigration between the groups of treated and control country pairs. The measures of online content are the bytes added and the number of edits to articles in the four analyzed knowledge domains (in rows). The time line is limited by 4 years before and after the shocks.

The dependent variable $WikiContent_{dot}$ encapsulates measures of contributions to Wikipedia on the languages of origin countries about host countries. These are the logarithm of the total number of edits, edits performed by unregistered (anonymous) users, by registered and logged-in users, the number of unregistered as well as registered users who worked on each knowledge domain (Research Centres of the host country, Software of the host country, and Cuisine of the host country) over the year.

3.2 Results

The results of OLS estimation of equation (1) are presented in Table (3). The independent variable of interest is the logarithm of the number of high-skilled immigrants from the origin countries who entered the host countries in each year. There is a number of dependent variables available for each knowledge domain, and in Table (3) I focus on the total number of edits, among them on edits performed by unregistered users and by those who signed up into Wikipedia with their user accounts, and also on users who appear in the data for the first time, i.e. new users, and returning users. All dependent variables are in logarithms as well. All specifications include year dummies and the control measure for the development of each Wikipedia language edition, which is the logarithm of the total number of articles in each language.

The results suggest that for all considered domains of Wikipedia, "Research institutes", "Software" as well as "Cuisine" of the host country more content is generated in the native languages of immigrants once the inflow of high-skilled immigrants into the host country increases. Moreover, for knowledge-intensive domains "Research institutes" and "Software" the magnitude of the effects is stronger. For "Research institutes" and "Cuisine" the increase in edits seems to be driven by an activity of anonymous editors. As some previous studies on Wikipedia suggest, editors, when they at first come to Wikipedia, might be skipping the registration procedure before contributing if they do that occasionally. After some time, when contributing becomes a more systematic activity for them, they register and make authorization before contributing. Therefore, occasional anonymous contributions could be more sensitive to immigration inflows. The number of observations for knowledge intensive domains is smaller than for "Cuisine" in all following specifications, which is due to the fact that those articles have lower readership and, hence, get fewer contributions and language coverage.

In order to ensure that the observed in OLS effects are causal, I test the DiD approach presented in equation (2). The DiD results (Table 4) support the main hypothesis for all three knowledge domains: contributions to Wikipedia are more intense between pairs of countries affected by the spikes in immigration in the years of the spike as well as after the spike. The strongest effects are observed for knowledge-intensive domains on research institutes and software. The results suggest that the increase in edits in the two knowledge intensive domains is driven to a greater extent by an increase in an anonymous editing activity.

Table 3: Skilled Immigration Flows and Content Generation on Wikipedia: Fixed Effects Estimation.

	Research Institutes				Software				Cuisine			
	(1) Edits	(2) EditsAnon	(3) EditsReg	(4) NewRegUs	(5) Edits	(6) EditsAnon	(7) EditsReg	(8) NewRegUs	(9) Edits	(10) EditsAnon	(11) EditsReg	(12) NewRegUs
High-skill. immigr.	0.180	0.308**	0.092	0.092	0.382*	0.260	0.320*	0.099	0.153**	0.194**	0.139*	0.136**
	(0.113)	(0.126)	(0.144)	(0.134)	(0.189)	(0.233)	(0.182)	(0.190)	(0.074)	(0.091)	(0.074)	(0.058)
Articles in wikilang.	0.080	0.185**	0.047	0.511***	0.777	0.496	0.870	0.930	0.586***	0.519***	0.570***	0.543***
	(0.094)	(0.091)	(0.102)	(0.103)	(0.844)	(0.615)	(0.766)	(0.720)	(0.090)	(0.124)	(0.089)	(0.088)
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean dep. Variable	3.23	1.62	3.04	1.98	3.44	2.33	3.09	2.46	3.91	2.29	3.68	2.54
Observations	312	312	312	312	240	240	240	240	1413	1413	1413	1413
# of country pairs	49	49	49	49	37	37	37	37	217	217	217	217
R^2	0.064	0.079	0.049	0.282	0.077	0.062	0.075	0.094	0.149	0.066	0.164	0.206

NOTE: This table contains estimation results for different measures of Wikipedia content (the number of edits, anonymous and registered edits, and new registered users) about host countries on the languages of origin countries. The results for different knowledge domains on Wikipedia are in columns: (1) - (4) Research Institutions in the host country, (5) - (8) Software in the host country, and (9) - (12) Cuisine of the host country. The independent variable of interest is the log number of skilled 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 host-origin country pair level. Significance stars denote: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 4: Difference-in-differences analysis of Immigration Inflow Effect on Content Generation on Wikipedia.

	Research Institutes				Software				Cuisine			
	(1) Edits	(2) EditsAnon	(3) EditsReg	(4) NewRegUs	(5) Edits	(6) EditsAnon	(7) EditsReg	(8) NewRegUs	(9) Edits	(10) EditsAnon	(11) EditsReg	(12) NewRegUs
Treated 2 years before	0.200	0.185	0.130	-0.001	0.169	0.171	0.174	0.127	0.226	0.145	0.250*	0.289***
	(1.39)	(0.51)	(1.10)	(-0.01)	(0.71)	(0.78)	(0.72)	(0.55)	(1.54)	(1.27)	(1.79)	(2.97)
Treated 1 years before	0.376**	-0.017	0.381**	-0.242	0.389	0.394	0.374	0.408*	0.137	0.286*	0.116	0.249***
	(2.16)	(-0.06)	(2.13)	(-0.76)	(1.20)	(1.12)	(1.35)	(1.70)	(0.93)	(1.88)	(0.87)	(2.71)
Treated 0 years after	0.421**	0.511**	0.328*	0.224	0.621*	0.587*	0.578*	0.517*	0.304**	0.280**	0.296**	0.287***
	(2.21)	(2.48)	(1.82)	(1.17)	(1.95)	(1.76)	(1.89)	(1.77)	(2.18)	(2.10)	(2.21)	(2.85)
Treated 1 years after	0.474**	0.286	0.452**	0.096	0.673*	0.653	0.642**	0.600*	0.131	0.178	0.160	0.267**
	(2.09)	(0.98)	(2.09)	(0.42)	(2.00)	(1.59)	(2.12)	(1.85)	(0.74)	(1.04)	(0.93)	(2.36)
Treated 2 years after	0.362	-0.248	0.392	-0.350	0.945***	1.077***	0.771**	0.444	0.288	0.342*	0.260	0.367***
	(1.41)	(-0.64)	(1.42)	(-1.29)	(3.01)	(2.70)	(2.36)	(1.35)	(1.51)	(1.68)	(1.43)	(2.67)
Treated 3 years after	-0.379	0.245	-0.491	0.475	1.333***	1.044***	1.301**	1.196***	0.450**	0.462**	0.413**	0.402***
	(-0.83)	(0.64)	(-1.02)	(1.28)	(2.75)	(3.02)	(2.63)	(2.99)	(2.29)	(2.05)	(2.16)	(2.64)
Articles in wikilang.	0.032	0.432*	-0.071	0.480***	0.852	0.759	0.780	0.914	0.645***	0.681***	0.585***	0.584***
	(0.19)	(1.72)	(-0.45)	(3.21)	(1.21)	(1.31)	(1.25)	(1.46)	(6.63)	(5.66)	(6.53)	(7.53)
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean dep. Variable	3.27	1.66	3.07	1.95	3.49	2.40	3.10	2.50	3.89	2.26	3.67	2.49
Observations	343	343	343	343	313	313	313	313	2421	2421	2421	2421
# of country pairs	62	62	62	62	60	60	60	60	476	476	476	476
R^2	0.060	0.082	0.065	0.242	0.085	0.084	0.076	0.144	0.157	0.075	0.161	0.109

NOTE: This table contains estimation results for different measures of Wikipedia content (the number of edits, anonymous and registered edits, and new registered users) about host countries on the languages of origin countries. The results for different knowledge domains on Wikipedia are in columns: (1) - (4) Research Institutions in the host country, (5) - (8) Software in the host country, and (9) - (12) Cuisine of the host country. The independent variables of interest are dummy variables representing interaction terms of the origin-host country pairs treated by the spikes in immigration with dummies for each year before as well as after the shocks. All specifications include year dummies. All standard errors (in parentheses) are clustered at the host-origin country pair level. Significance stars denote: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

As a robustness check, I further re-define treated country pairs using only skilled bilateral immigrant flows (then, the estimation is performed only for those country pairs where the education shares are available). Focusing on spikes in high-skilled immigration inflows strengthens the magnitudes of the found positive effects of immigration on generation of meta knowledge about science and technology. However, the effects are lower for the domain "Cuisine", suggesting that contributions to this category are made to a lesser extent by highly skilled immigrants (see Table 5).

Table 5: Difference-in-differences analysis of Immigration Inflow Effect on Content Generation on Wikipedia.

	Research Institutes				Software				Cuisine			
	(1) Edits	(2) EditsAnon	(3) EditsReg	(4) NewRegUs	(5) Edits	(6) EditsAnon	(7) EditsReg	(8) NewRegUs	(9) Edits	(10) EditsAnon	(11) EditsReg	(12) NewRegUs
Treated 2 years before	-0.017	0.412	-0.134	-0.018	-0.124	-0.249	-0.027	-0.167	0.140	0.101	0.170	0.106
	(-0.11)	(1.61)	(-0.84)	(-0.16)	(-0.53)	(-1.02)	(-0.11)	(-0.88)	(0.76)	(0.48)	(0.97)	(0.71)
Treated 1 years before	0.114	0.306	0.036	-0.135	0.321	0.176	0.410	0.218	0.149	0.140	0.178	0.276***
	(0.73)	(1.28)	(0.21)	(-0.55)	(0.86)	(0.53)	(1.13)	(0.89)	(0.88)	(0.66)	(1.06)	(2.63)
Treated 0 years after	0.242	0.730***	0.106	0.138	0.473	0.533*	0.419	0.331	0.161	0.166	0.146	0.101
	(1.31)	(3.77)	(0.56)	(0.69)	(1.21)	(1.76)	(1.10)	(1.15)	(0.84)	(0.82)	(0.74)	(0.87)
Treated 1 years after	0.411	0.640**	0.261	0.042	0.363	0.438	0.358	0.096	0.054	0.222	0.062	0.150
	(1.62)	(2.22)	(1.04)	(0.18)	(0.83)	(1.06)	(0.79)	(0.30)	(0.19)	(0.93)	(0.23)	(0.96)
Treated 2 years after	0.223	0.340	0.031	-0.136	0.496	0.876	0.279	-0.053	0.259	0.249	0.296	0.077
	(0.88)	(0.97)	(0.10)	(-0.54)	(0.99)	(1.60)	(0.61)	(-0.13)	(0.90)	(0.85)	(1.04)	(0.42)
Treated 3 years after	-0.247	0.485	-0.460	0.152	1.201	1.254**	0.876	-0.080	0.464	0.299	0.528	0.030
	(-0.57)	(0.92)	(-1.06)	(0.35)	(1.44)	(2.08)	(0.92)	(-0.12)	(1.43)	(0.86)	(1.64)	(0.15)
Articles in wikilang.	0.134	0.255*	0.065	0.455***	0.608	0.388	0.681	0.700	0.638***	0.619***	0.600***	0.633***
	(0.91)	(1.86)	(0.46)	(3.40)	(0.64)	(0.53)	(0.83)	(0.84)	(5.87)	(4.13)	(5.47)	(7.37)
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean dep. Variable	3.27	1.64	3.08	1.96	3.53	2.39	3.17	2.52	3.91	2.30	3.69	2.49
Observations	264	264	264	264	197	197	197	197	1176	1176	1176	1176
# of country pairs	46	46	46	46	36	36	36	36	212	212	212	212
R^2	0.079	0.089	0.067	0.283	0.071	0.092	0.069	0.113	0.152	0.063	0.164	0.112

NOTE: This table contains estimation results for different measures of Wikipedia content (the number of edits, anonymous and registered edits, and new registered users) about host countries on the languages of origin countries. The results for different knowledge domains on Wikipedia are in columns: (1) - (4) Research Institutions in the host country, (5) - (8) Software in the host country, and (9) - (12) Cuisine of the host country. The independent variables of interest are dummy variables representing interaction terms of the origin-host country pairs treated by the spikes in immigration with dummies for each year before as well as after the shocks. All specifications include year dummies. All standard errors (in parentheses) are clustered at the host-origin country pair level. Significance stars denote: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

4 Inverse Spillovers: Does Knowledge About Origin Countries Disseminate in Host Countries?

In this Section I test whether the found knowledge spillovers are one- or bi-directional. If, in addition to direct, reverse spillovers were present, then knowledge about origin countries' science, institutions and cuisine would become more available on the languages of host countries. To examine the reverse spillover, I estimate the following regression equation:

$$Content_{odt} = \alpha\ ImmigrationInflow_{dot} + \beta\ Controls_{ot} + \gamma_{do} + \delta_t + \epsilon_{dot}, \tag{3}$$

where now $Content_{odt}$ is Wikipedia content contributed about the origin country o in the language of the destination country d due to immigration to country d from country o, t is the current year. As before, time dummies, country pair fixed effects are included as well as the total number of articles in the corresponding language edition. The estimation results of equation (3) are presented in Table (6). They suggest no evidence for the reverse spillovers from the inflow of immigrants to host countries. For example, when the inflow of immigrants from Russia moves to Germany, there is no evidence that more information about Russia is contributed in German language.

Table 6: Immigration Flows and Content Generation on Wikipedia: Fixed Effects Estimation.

	Research Institutes				Software				Cuisine			
	(1) Edits	(2) EditsAnon	(3) EditsReg	(4) NewRegUs	(5) Edits	(6) EditsAnon	(7) EditsReg	(8) NewRegUs	(9) Edits	(10) EditsAnon	(11) EditsReg	(12) NewRegUs
Immigrants	-0.049	0.018	-0.063	-0.087	-0.106	-0.113	-0.137	-0.132	-0.016	-0.003	-0.023	-0.014
	(0.156)	(0.136)	(0.147)	(0.099)	(0.126)	(0.225)	(0.083)	(0.149)	(0.022)	(0.028)	(0.021)	(0.017)
Articles in wikilang.	0.084	0.345*	0.035	0.776***	1.372***	1.254*	1.205***	1.266**	0.421***	0.559***	0.332***	0.524***
	(0.186)	(0.202)	(0.191)	(0.136)	(0.445)	(0.663)	(0.376)	(0.523)	(0.069)	(0.094)	(0.066)	(0.061)
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean dep. Variable	3.20	1.74	2.98	2.00	4.05	3.01	3.60	3.04	4.03	2.57	3.76	2.77
Observations	459	459	459	459	338	338	338	338	4412	4412	4412	4412
# of country pairs	72	72	72	72	48	48	48	48	640	640	640	640
R^2	0.055	0.089	0.062	0.160	0.149	0.120	0.117	0.121	0.121	0.083	0.135	0.156

NOTE: This table contains estimation results for different measures of Wikipedia content (the number of edits, anonymous and registered edits, and new registered users) about host countries on the languages of origin countries. The results for different knowledge domains on Wikipedia are in columns: (1) - (4) Research Institutions in the host country, (5) - (8) Software in the host country, and (9) - (12) Cuisine of the host country. 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 host-origin country pair level. Significance stars denote: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Analogically to the benchmark analysis, the equation for the main DiD specification (equation 2) is reformulated for the inverse spillovers. The results are in Table (7).

Table 7: The Inverse Spillovers: Difference-in-differences analysis.

	Research Institutes				Software				Cuisine			
	(1) Edits	(2) EditsAnon	(3) EditsReg	(4) NewRegUs	(5) Edits	(6) EditsAnon	(7) EditsReg	(8) NewRegUs	(9) Edits	(10) EditsAnon	(11) EditsReg	(12) NewRegUs
Treated 2 years before	-0.264	-0.216	-0.229	0.129	0.093	0.170	0.068	0.091	-0.005	-0.030	0.008	0.029
	(-1.34)	(-0.64)	(-1.26)	(0.53)	(0.28)	(0.50)	(0.20)	(0.37)	(-0.07)	(-0.35)	(0.11)	(0.45)
Treated 1 years before	-0.418*	-0.230	-0.380	0.044	-0.156	0.007	-0.156	-0.058	-0.136	-0.103	-0.128	-0.117
	(-1.75)	(-0.88)	(-1.65)	(0.14)	(-0.51)	(0.02)	(-0.53)	(-0.23)	(-1.48)	(-1.06)	(-1.45)	(-1.50)
Treated 0 years after	-0.348	-0.045	-0.353	0.100	0.158	-0.072	0.254	0.169	-0.163	-0.130	-0.123	-0.039
	(-1.18)	(-0.14)	(-1.32)	(0.35)	(0.37)	(-0.14)	(0.65)	(0.48)	(-1.63)	(-1.25)	(-1.26)	(-0.49)
Treated 1 years after	-0.055	-0.142	-0.034	0.132	0.023	-0.068	0.019	0.164	-0.124	-0.217**	-0.070	-0.033
	(-0.23)	(-0.57)	(-0.14)	(0.55)	(0.07)	(-0.12)	(0.07)	(0.43)	(-1.16)	(-2.16)	(-0.65)	(-0.38)
Treated 2 years after	-0.336	-0.385	-0.266	-0.020	-0.246	-0.488	-0.152	-0.057	-0.141	-0.203*	-0.095	-0.089
	(-1.36)	(-1.48)	(-1.06)	(-0.12)	(-0.61)	(-0.78)	(-0.41)	(-0.13)	(-1.18)	(-1.74)	(-0.79)	(-1.00)
Treated 3 years after	0.260	-0.375	0.389	0.216	-0.268	-0.169	-0.156	0.028	-0.177	-0.177	-0.143	-0.052
	(0.85)	(-1.38)	(1.18)	(0.70)	(-0.59)	(-0.28)	(-0.41)	(0.06)	(-1.23)	(-1.23)	(-1.03)	(-0.48)
Articles in wikilang.	0.211	0.669*	0.064	0.867***	1.698***	1.387**	1.626***	1.246**	0.616***	1.018***	0.448***	0.737***
	(0.61)	(1.96)	(0.19)	(3.79)	(4.17)	(2.27)	(4.64)	(2.50)	(5.93)	(8.12)	(4.30)	(8.01)
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean dep. Variable	3.19	1.70	2.98	1.96	4.03	2.96	3.60	3.03	4.03	2.53	3.78	2.72
Observations	377	377	377	377	280	280	280	280	3626	3626	3626	3626
# of country pairs	69	69	69	69	47	47	47	47	637	637	637	637
R^2	0.073	0.088	0.078	0.165	0.170	0.086	0.157	0.116	0.146	0.100	0.144	0.124

NOTE: This table contains estimation results for different measures of Wikipedia content (the number of edits, anonymous and registered edits, and new registered users) about host countries on the languages of origin countries. The results for different knowledge domains on Wikipedia are in columns: (1) - (4) Research Institutions in the host country, (5) - (8) Software in the host country, and (9) - (12) Cuisine of the host country. The independent variables of interest are dummy variables representing interaction terms of the origin-host country pairs treated by the spikes in immigration with dummies for each year before as well as after the shocks. All specifications include year dummies. All standard errors (in parentheses) are clustered at the host-origin country pair level. Significance stars denote: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

5 Robustness checks

To ensure that the results are robust and reflect the true effects of immigration on content availability on Wikipedia I perform several robustness checks. I check, whether the choice of the shock moment for the origin-host country pairs used as a control group matters for the result, and it turns that the results hold independently of the choice of that moment in 2008-2012 (see Appendix, Table 9).

I also ensure that the results I find are caused by the difference between large and modest immigrant inflows to host countries, and are not just capturing the overall growth of Wikipedia in particular, less represented languages over the last five years. Among the control group, I create a placebo treatment group, randomly assigning origin-host country into new treated and control groups. Then, I estimate the same difference-in-differences model as in the main Section. The results show the absence of effects of interest (see Appendix, Table 10).

6 Wikipedia Content and Student Mobility

6.0.1 OLS Regression

In order to study some potential implications of online information freely available on Wikipedia I test the relationship between the student choices of countries for studying abroad and content generated on Wikipedia in the domains of science and research institutions in the previous year in the panel data OLS specification:

$$Nstudents_{dot} = \alpha_{do} + \beta\ WikiContent_{dot} + \nu_t + \epsilon_{dot}, \tag{4}$$

where d stands for the destination (host) country or the topic dedicated to it on Wikipedia, o for the origin country or its language for Wikipedia content, t is the current year. Country pair fixed effects are included to rule out time-invariant unobserved heterogeneity and time fixed effects to control for common the time trend. All dependent variables are in logarithms.

6.1 Results

The estimation results of equation (4) are presented in Table (8). The results suggest that changes in total edits of Wikipedia content on the topics related to science and technology are positively related to next year's choices of students on where to spend abroad semesters. For other host country-specific knowledge domains the results turn insignificant.

Table 8: Content on Wikipedia about Science and Technology of the Destination Countries and Student Mobility Choices: Fixed Effects Estimation.

	International Student Mobility			
	(1)	(2)	(3)	(4)
Science and tech. (edits), t	0.008			
	(0.27)			
Science and tech. (edits), t-1		0.052***		
		(2.70)		
Science and tech. (bytes), t			0.013	
			(0.94)	
Science and tech. (bytes), t-1				0.032***
				(3.20)
Year dummies	Yes	Yes	Yes	Yes
Mean dep. Variable	4.97	4.99	4.97	4.99
Observations	2305	2167	2305	2167
# of country pairs	448	443	448	443
R^2	0.206	0.203	0.206	0.207

NOTE: This table contains the OLS estimates for the relationship between the changes in content available on Wikipedia for domain , measured by edits and bytes, and the numnber of international students chosing the destination countries for their abroad semesters. The independent variable of interest is the log number content generated on the language of origin country *o* about destination country *d* in years *t* and *t-1*. The dependent variable is student mobility, measured as a log number of students from origin country *o* arriving for studies to destination country *d* in year *t*. All specifications include year dummies. All standard errors (in parentheses) are clustered at the host-origin country pair level. Significance stars denote: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

7 Conclusion

This paper sheds light on the potential impact of immigration on dissemination of meta knowledge. It uses high spikes in immigration to identify the causal effect of immigration inflows on contributions to Wikipedia about host countries in languages of immigrants' origin countries. The nature of Wikipedia allows investigating its content development using a number of indicators, in particular, I analyze edits performed by registered or anonymous (unregistered) contributors, bytes of information, the number of registered contributors and the number of articles created over each period.

My findings suggest that more online knowledge becomes available about the host countries on the languages of origin countries on Wikipedia when spikes in immigration occur. For knowledge-intensive domains "Research institutes" and "Software" the magnitude of the effects is stronger, and an increase in the contributing activity is driven by anonymous editors, who in line with Wikipedia philosophy could be skipping the registration procedure because they are either inexperienced contributors or occasional contributors. This results are robust in several checks and to restricting immigration flows to considering only immigrants with at least higher education.

At the same time, I find no evidence of the inverse spillovers, i.e. for more content contributed about the origin countries in the languages of the host countries.

This knowledge might have wide implications for the technology adoption, individual knowledge-

related choices and, therefore, development paths in the immigrants' origin countries. In particular, I find that more information available about science and technology and research centres of the host countries attracts more students from the origin country in the subsequent years. However, a substantial further analysis is needed to evaluate the extent to which source countries could benefit from the meta knowledge about technology related topics, for example, in terms of impact on consumer choices, technology adoption and entrepreneurship opportunities.

The evidence from this study also highlights an important role of information and communication technologies, which became more widely available in the last decade, in cross-border knowledge facilitation and mitigation of the negative effects of brain drain for poorer countries.

References

Asatryan, Zareh, Benjamin Bittschi, and Philipp Doerrenberg, "Remittances and public finances: Evidence from oil-price shocks," *ZEW-Centre for European Economic Research Discussion Paper*, 2016, (16-022).

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.

Bruecker, Herbert, Stella Capuano, and Abdeslam Marfouk, "Education, gender and international migration: Insights from a panel dataset 1980-2010–Methodology Report," *mimeo Institute for Employment Research*, 2013.

Chiquiar, Daniel and Gordon H Hanson, "International migration, self-selection, and the distribution of wages: Evidence from Mexico and the United States," 2002.

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.

Foley, C Fritz and William R Kerr, "Ethnic innovation and US multinational firm activity," *Management Science*, 2013, *59* (7), 1529–1544.

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.

Grogger, Jeffrey and Gordon H Hanson, "Income maximization and the selection and sorting of international migrants," *Journal of Development Economics*, 2011, *95* (1), 42–57.

Hergueux, Jérôme, Yann Algan, Yochai Benkler, and Mayo Fuster Morell, "Cooperation in Peer Production Economy: Experimental Evidence from Wikipedia," in "Lyon Meeting" 2014.

Hunt, Jennifer and Marjolaine Gauthier-Loiselle, "How much does immigration boost innovation?," *American Economic Journal: Macroeconomics*, 2010, *2* (2), 31–56.

Kummer, Michael E, "Spillovers in Networks of User Generated Content," *Available at SSRN*, 2013.

Ozgen, Ceren, Cornelius Peters, Annekatrin Niebuhr, Peter Nijkamp, and Jacques Poot, "Does Cultural Diversity of Migrant Employees Affect Innovation?," *International Migration Review*, 2014, *48* (s1), S377–S416.

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).

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

Table 9: Difference-in-differences analysis of Immigration Inflow Effect on Content Generation on Wikipedia

	Research Institutes				Software				Cuisine			
	(1) Edits	(2) EditsAnon	(3) EditsReg	(4) NewRegUs	(5) Edits	(6) EditsAnon	(7) EditsReg	(8) NewRegUs	(9) Edits	(10) EditsAnon	(11) EditsReg	(12) NewRegUs
Treated 2 years before	0.205	0.308	0.115	0.037	-0.093	-0.061	-0.062	-0.031	0.187	0.144	0.209	0.246**
	(1.39)	(0.94)	(0.87)	(0.26)	(-0.34)	(-0.28)	(-0.21)	(-0.12)	(1.28)	(1.32)	(1.50)	(2.44)
Treated 1 years before	0.384**	0.125	0.367*	-0.205	0.107	0.130	0.122	0.198	0.181	0.310**	0.172	0.233**
	(2.05)	(0.56)	(1.92)	(-0.64)	(0.40)	(0.44)	(0.52)	(1.04)	(1.24)	(2.08)	(1.30)	(2.42)
Treated 0 years after	0.459**	0.687***	0.347**	0.315	0.115	0.107	0.127	0.144	0.353***	0.318***	0.365***	0.200*
	(2.43)	(2.72)	(2.12)	(1.38)	(0.39)	(0.36)	(0.40)	(0.54)	(2.63)	(2.66)	(2.76)	(1.90)
Treated 1 years after	0.514***	0.476**	0.477***	0.231	0.030	0.025	0.079	0.135	0.195	0.233	0.246	0.168
	(2.76)	(2.15)	(2.73)	(1.08)	(0.10)	(0.07)	(0.26)	(0.46)	(1.20)	(1.58)	(1.51)	(1.54)
Treated 2 years after	0.422***	0.041	0.432***	-0.081	0.127	0.271	0.054	-0.145	0.357**	0.391**	0.361**	0.251**
	(2.78)	(0.17)	(2.89)	(-0.36)	(0.43)	(0.62)	(0.16)	(-0.52)	(2.14)	(2.32)	(2.22)	(2.01)
Treated 3 years after	0.079	0.516**	0.010	0.749**	0.530	0.173	0.595	0.474	0.524***	0.506***	0.530***	0.268**
	(0.20)	(2.46)	(0.02)	(2.33)	(1.10)	(0.36)	(1.31)	(1.45)	(3.15)	(2.87)	(3.20)	(2.15)
Articles in wikilang.	-0.066	-0.071	-0.032	0.646***	1.402	1.300*	1.362	1.468*	0.557***	0.460***	0.526***	0.463***
	(-0.24)	(-0.29)	(-0.12)	(3.18)	(1.32)	(1.69)	(1.36)	(1.84)	(5.72)	(4.69)	(5.54)	(6.34)
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean dep. Variable	3.20	1.62	3.01	1.86	3.37	2.31	2.98	2.36	3.88	2.26	3.65	2.47
Observations	334	334	334	334	309	309	309	309	2383	2383	2383	2383
# of country pairs	65	65	65	65	61	61	61	61	484	484	484	484
R^2	0.051	0.061	0.058	0.247	0.116	0.077	0.109	0.144	0.090	0.037	0.107	0.201

NOTE: This table contains estimation results for different measures of Wikipedia content (the number of edits, anonymous and registered edits, and new registered users) about host countries on the languages of origin countries. The results for different knowledge domains on Wikipedia are in columns: (1) - (4) Research Institutions in the host country, (5) - (8) Software in the host country, and (9) - (12) Cuisine of the host country. The independent variables of interest are dummy variables representing interaction terms of the origin-host country pairs treated by the spikes in immigration with dummies for each year before as well as after the shocks. All specifications include year dummies. All standard errors (in parentheses) are clustered at the host-origin country pair level. Significance stars denote: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 10: Difference-in-differences analysis of Immigration Inflow Effect on Content Generation on Wikipedia

	Research Institutes				Software				Cuisine			
	(1) Edits	(2) EditsAnon	(3) EditsReg	(4) NewRegUs	(5) Edits	(6) EditsAnon	(7) EditsReg	(8) NewRegUs	(9) Edits	(10) EditsAnon	(11) EditsReg	(12) NewRegUs
Treated 2 years before	-0.693*	-0.291	-0.683*	-0.482*	-0.338	-0.320	-0.228	-0.381	0.051	0.204*	-0.037	0.153*
	(-1.99)	(-0.94)	(-1.83)	(-1.75)	(-1.02)	(-0.66)	(-0.79)	(-1.66)	(0.52)	(1.97)	(-0.35)	(1.84)
Treated 1 years before	-0.415	-0.518	-0.305	-0.392	-0.705	-0.856*	-0.520	-0.683*	0.096	0.201*	-0.001	0.139*
	(-1.34)	(-1.43)	(-1.00)	(-1.64)	(-1.58)	(-1.84)	(-1.21)	(-1.86)	(0.94)	(1.74)	(-0.01)	(1.68)
Treated 0 years after	-0.190	-0.122	-0.177	-0.187	-0.518	-0.492	-0.407	-0.677*	0.046	0.229**	-0.055	0.107
	(-0.53)	(-0.31)	(-0.49)	(-0.63)	(-1.11)	(-0.98)	(-0.94)	(-1.84)	(0.45)	(2.14)	(-0.49)	(1.25)
Treated 1 years after	-0.162	-0.248	-0.051	-0.261	-0.392	-0.123	-0.385	-0.324	-0.105	0.068	-0.152	0.030
	(-0.52)	(-0.73)	(-0.16)	(-1.07)	(-0.79)	(-0.23)	(-0.85)	(-0.82)	(-0.96)	(0.59)	(-1.30)	(0.31)
Treated 2 years after	-0.267	-0.430	-0.245	-0.299	-0.439	-0.525	-0.243	-0.405	0.004	0.084	-0.039	0.081
	(-0.69)	(-1.35)	(-0.62)	(-1.34)	(-0.88)	(-0.94)	(-0.52)	(-0.92)	(0.04)	(0.75)	(-0.32)	(0.86)
Treated 3 years after	-0.591*	-0.206	-0.563	-0.397	-0.396	-0.586	-0.360	-0.349	-0.052	-0.024	-0.075	0.029
	(-1.74)	(-0.59)	(-1.66)	(-1.53)	(-0.85)	(-1.18)	(-0.82)	(-0.86)	(-0.45)	(-0.19)	(-0.63)	(0.33)
Articles in wikilang.	0.033	0.497*	-0.083	0.516***	0.931	0.797	0.870	0.995	0.580***	0.545***	0.553***	0.525***
	(0.20)	(1.80)	(-0.55)	(3.26)	(1.29)	(1.36)	(1.38)	(1.53)	(6.71)	(5.71)	(6.06)	(8.37)
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean dep. Variable	3.24	1.64	3.04	1.94	3.41	2.33	3.02	2.45	3.85	2.23	3.62	2.45
Observations	284	284	284	284	242	242	242	242	2207	2207	2207	2207
# of country pairs	49	49	49	49	43	43	43	43	386	386	386	386
R^2	0.070	0.083	0.073	0.236	0.077	0.098	0.059	0.169	0.152	0.066	0.159	0.094

NOTE: This table contains estimation results for different measures of Wikipedia content (the number of edits, anonymous and registered edits, and new registered users) about host countries on the languages of origin countries. The results for different knowledge domains on Wikipedia are in columns: (1) - (4) Research Institutions in the host country, (5) - (8) Software in the host country, and (9) - (12) Cuisine of the host country. The independent variables of interest are dummy variables representing interaction terms of the origin-host country pairs treated by the spikes in immigration with dummies for each year before as well as after the shocks. All specifications include year dummies. All standard errors (in parentheses) are clustered at the host-origin country pair level. Significance stars denote: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.