CESifo WORKING PAPERS

7569
2019

March 2019

Presence of language-learning opportunities and migration

Matthias Huber, Silke Uebelmesser

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
www.cesifo-group.org/wp

An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: www.CESifo-group.org/wp

CESifo *Working Paper No. 7569*
Category 4: Labour Markets

Presence of language-learning opportunities and migration

Abstract

This paper analyses the effect of the presence of German language learning opportunities abroad on migration to Germany. We use information on the presence of the Goethe-Institut (GI), which is an association that promotes German culture and offers language courses and standardized exams. Our unique dataset covers 69 countries for the period 1977 to 2014. In this multiple-origin and single-destination framework, we estimate fixed-effects models as our basic specification. We find evidence that the number of language institutes of the GI in a country is positively correlated with migration from that country to Germany. The correlation is higher for countries with lower income, larger linguistic distance and without wars. To establish causality, we consider Switzerland as an alternative destination country as the decision to open a language institute in a country is exogenous to migration flows from that country to Switzerland. We find that the institutes of the GI also affect migration flows to the German-speaking part of Switzerland, but not to the French- and Italian-speaking part. Backed by further extensions which control for the presence of multilateral resistance and omitted variable bias, we interpret our results as presenting a causal effect from language learning opportunities to migration flows.

JEL-Codes: F220, O150, J610.

Keywords: language skills, language learning, international migration, panel data.

Matthias Huber
University of Jena
Carl-Zeiss-Str. 3
Germany – 07743 Jena
matthias.huber@uni-jena.de

Silke Uebelmesser
University of Jena
Carl-Zeiss-Str. 3
Germany – 07743 Jena
silke.uebelmesser@uni-jena.de

March 21, 2019
We are grateful for comments and suggestions from participants of the CEMIR/ifo Workshop, the CESifo Area Conference Employment and Social Protection, the European Association of Labour Economists (EALE) Conference, the Dondena Workshop on Public Policy at Bocconi, the CReAm/RWI Workshop on the Economics of Migration in 2017, and the European Economic Association Congress, the OECD International Forum on Migration Statistics and the Spring Meeting of Young Economists (SMYE) in 2018. This work was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation, UE 124/2-1) – 270886786.

1 Introduction

A large part of the migration literature focuses on migrants' proficiency in the language of the destination country. It has been shown that proficiency improves labour market and integration outcomes. Language skills increase earnings (see e.g. Dustmann and Soest 2001; Chiswick and Miller 1995) and employment probability (Dustmann and Fabbri 2003). At the same time, the probability of intermarriage becomes larger and the likelihood of living in an ethnic enclave decreases (Bleakley and Chin 2010).

Given these benefits of language proficiency in the destination country, potential migrants can be expected to consider languages in their migration decision and their location choice. Indeed, many studies have shown that language is an important determinant of migration flows. For this, measures of linguistic distance are often used to capture the linguistic relationship between the migrants' mother tongue and the language of the destination country. Adserà and Pytliková (2015) and Belot and Hatton (2012) find evidence of a negative effect of linguistic distance on international migration flows based on different measures. However, the concept of linguistic distance neglects language acquisition of potential migrants before migration which can alleviate or overcome the negative effects of linguistic distance.

The aim of this paper is to study the effect of the presence of language learning opportunities on migration. We build a random-utility model. Individuals want to maximize expected utility of migration, of which the expected wage income net of migration costs is an important component. Acquiring language skills of the host country can increase expected net wage income if the benefits, in terms of higher wages abroad, exceed the costs of learning. Language acquisition can happen at different points in time. We distinguish language learning in earlier years and as an adult, as this is different in terms of costs and direction of causality in the context of migration. If language skills are acquired during childhood or adolescence, the decision is more likely determined by factors outside the learners' direct control. These factors may be related to the school system with compulsory foreign language learning and to parents' preferences. Language proficiency often comes with no or little costs, or they can be regarded as sunk costs, while it might affect later migration decisions. Aparicio Fenoll and Kuehn (2016) use compulsory language learning at school as a measure for language skills beyond linguistic properties and find a positive relationship with migration flows within the European Union.

Language learning during adulthood, on the contrary, is more likely to be a decision of the learners themselves and driven by different motives which can be of personal or economic nature. Economic motives can relate to better job opportunities at home or abroad. Migration decisions might lead to pre-migration language learning. The direction of causality with language learning of adults can thus be opposite to the direction with language learning during childhood or adolescence. Uebelmesser and Weingarten (2017) analyse determinants of language learning of adults by using data from language institutes worldwide and in Germany. They show that general migration and student migration are indeed important determinants of language learning.

For our paper, we focus on language learning of adults. We use a unique panel dataset for 69 countries for the period 1977 to 2014 with information about the worldwide presence of

Goethe institutes collected from the annual reports of the Goethe-Institut (GI).[1] The GI is an association which is one of the main actors in promoting German culture and language worldwide. Via its institutes, it offers language courses and standardized language exams as well as information on German culture and society in many different forms, such as cultural events and libraries.

We find evidence that the number of language institutes in a country is positively correlated with migration from that country to Germany. By distinguishing between institutes that offer language services and those which do not, we can show that the correlation is indeed driven by language learning opportunities and not by other factors, such as the provision of information about German culture and society. The relation is stronger for poorer countries and for countries that are not involved in a war. Furthermore, we find that the number of language institutes has a larger effect on migration for countries where the majority does not speak a Germanic language.

For a causal interpretation, the issue of reverse causality has to be addressed. The GI and the Federal Foreign Office (FFO) jointly decide where to open and close institutes in a complex process.[2] Even though there are no indications that migration to Germany plays any particular role in this process, a possible impact cannot be fully ruled out. Therefore, we also estimate migration flows to the German-speaking part of Switzerland. As migration flows to Switzerland do not have any impact on the decision to open or close institutes, the number of institutes can be considered exogenous to those migration flows. With this specification, we find similar results for the effect of the number of institutes on migration flows to Switzerland as to Germany. Backed by further extensions which control for the presence of multilateral resistance (Bertoli and Fernández-Huertas Moraga 2013) and omitted variable bias, we conclude that the effect of the number of language institutes on migration can be interpreted in a causal way.

Our paper contributes to the literature in several ways: First, our unique data set provides new information on language learning for a long time period of 38 years and a large number of countries all over the world. Second, this data allows us to study the language learning process for adults in the context of migration in a comprehensive way. So far, the literature mainly looked at linguistic properties. While the only study that deals with language learning studies this during childhood (Aparicio Fenoll and Kuehn 2016), we investigate language learning of adults who can be expected to decide about this investment themselves.

The remainder of the paper is structured as follows: Section 2 describes the history of the GI, its institutional framework and its objectives. Furthermore, it explains the decision process of opening and closing institutes. Section 3 presents the data and provides descriptive statistics. Section 4 introduces a random utility maximization (RUM) model for the individual migration decisions and derives the estimation strategy. Section 5 presents our main results. In Section 6, we provide evidence to support a causal interpretation and deal with potential issues that threaten exogeneity. Section 7 concludes.

[1] In this paper we stick to the following convention: when referring to the association of the Goethe-Institut we use the abbreviation "GI". When talking about specific branches of the GI abroad, we refer to them as "institutes".

[2] See Section 2 for more information about the institutional background.

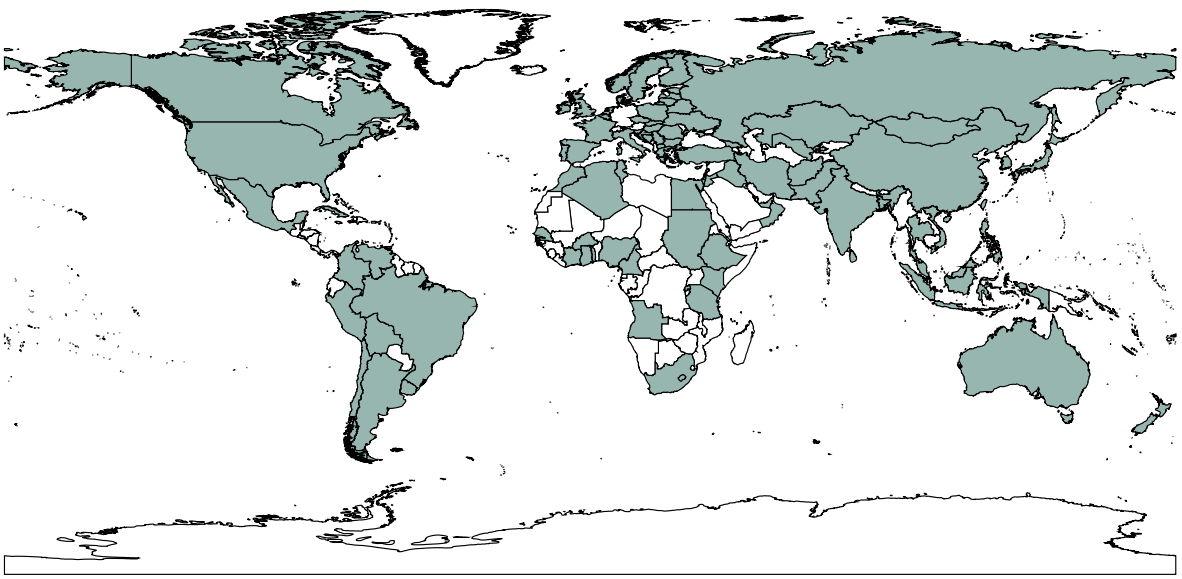


Figure 1: The distribution of the GI in 2014.

2 The Goethe-Institut

The GI is a German association which promotes German language and culture. It is one of the main actors of the German foreign cultural policy and therefore it is closely related to the German government, in particular to the Federal Foreign Office (FFO). The GI was founded in 1951 as a successor of the "Deutsche Akademie" (German academy) to rebuild and renew the infrastructure for foreign cultural policy after the Second World War. One of the main duties of the GI is to promote the German language. It does so by providing language courses and standardized exams, as well as scholarships for language learning and training for teachers. Beside language learning, it supports international cultural cooperation and provides information about the social, political and cultural life in Germany. For these reasons, the GI maintains libraries and organizes cultural events (Goethe-Institute and Auswärtiges Amt 2004). While the GI is mainly funded by the German government, language courses are financed by course fees (Goethe-Institut 2014).

To fulfill its duties, the GI has institutes all over the world. In 2014, there were in total 143 institutes in 93 countries (see Figure 1) out of which 126 institutes offered language services. The sizes of those institutes in terms of registrations for language courses ranged from 2 registrations in Beirut to 6800 registrations in Bangkok. The average number of registrations amounted to 1963. In total, there were 229,702 course registrations in 2014.

The process of opening and closing institutes is of particular importance for the analysis of this paper. According to officials of the GI, the FFO and the GI make this decision together. The process of opening new institutes starts with suggestions for locations by the GI, the FFO or members of the legislature. In a next step, the GI and the FFO discuss the potential locations on the basis of their general objectives, while at the same time considering the appropriate institutional type.[3] This complex process takes into account legal, political and social aspects of the potential host country. Furthermore, aspects related to the overall security in the host country and global developments play a role, as

[3]In this paper, we focus on institutes as this is the main type and abstract from any other types.

well as considerations about the region of the host country. Very often location decisions can be seen as reactions to changes in the political, social or economic situation. The most important example is the expansion of the GI to former socialist countries after the fall of the iron curtain. However, it cannot be ruled out that migration plays an indirect role in the location decision, even though the GI assures that opening an institute does not aim at creating brain drain in the host countries.[4]

3 Data and descriptive statistics

In the following, we describe the data used for the analysis and present descriptive statistics.

3.1 Dependent variable: migration rate

For the dependent variable, we use the yearly migration rate, i.e. immigration flows to Germany divided by the population size of the origin country. Migration data is provided in the 'Wanderungsstatistik' by the German Federal Statistical Office (Destatis 2016). The data documents the number of foreign citizens that move to Germany and register their residence in a given year. These immigrants are then categorized according to their citizenship. As this registration is mandatory for all foreign residents staying for more than 2 months, this data represents legal immigration to Germany in a comprehensive way. Data on population size comes from the Penn World Table (PWT) 9.0 (Feenstra et al. 2015).

3.2 Independent variable: language learning opportunities

The data on language learning opportunities is derived from a new dataset comprising information about the presence of the GI and the number of institutes at the country-level (Uebelmesser et al. 2018b). The GI has published annual reports continuously since 1965 in which activities of each institute, including statistics of language course and exam participation, have been reported. The dataset is constructed from these reports and contains information about the presence of GI on the city-level (see Uebelmesser et al. 2018a for a more detailed description of the dataset).

For our analysis, we aggregate the data on the number of institutes in a given country and year. As not every institute offers language services, we construct three different variables: the number of institutes without language services, the number of language institutes, i.e. institutes that offer language services, and the number of all institutes, i.e. the sum of the number of institutes with and without language services. As we want to show the effect of language learning opportunities, for our basic specification we restrict attention to institutes with language services. For robustness checks, we also use information about institutes without language services.

3.3 Other control variables

Several further variables are included as control variables. First, we control for the economic condition in the origin country by including GDP per capita. For this, we use the expenditure-side real GDP ('rgdpe') and data on the population size both from the

[4]We deal with the issue of potential reverse causality in Section 6.

Table 1: Summary statistics

Variable	Obs		Mean	Std. Dev.	Min	Max
Migration rate	2622	Overall	.0002	.0007	5.48e-07	.0111
All institutes	2622	Overall	.8955	1.3984	0	12
		Between		1.3237	0	8.4474
		Within		.4776	-3.5519	4.7902
Language institutes	2622	Overall	.8066	1.1587	0	9
		Between		1.1030	0	6.5
		Within		.3783	-1.7986	4.7014
Language institutes/population (in m)	2622	Overall	.0608	.1071	0	.6316
GDP per capita	2622	Overall	8904.7410	11076.0200	142.3924	74113.95
EU member	2622	Overall	.0587	.2352	0	1
Population (in m)	2622	Overall	39.0150	122.3834	.2221	1295.2920
Conflict	2622	Overall	.1690	.3748	0	1
War	2622	Overall	.0667	.2496	0	1
Migrant stock/population	2622	Overall	.0013	.0039	7.83e-06	.0349
Exports + imports (in EUR)	2622	Overall	3.90e+09	1.18e+10	75000	1.45e+11

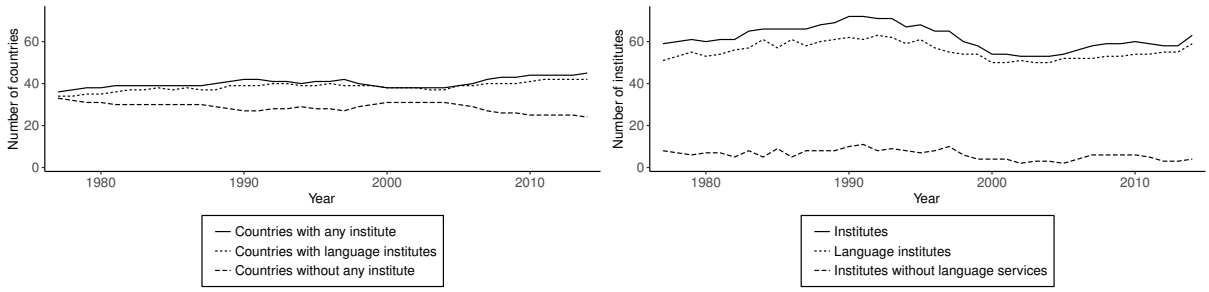
PWT 9.0 (Feenstra et al. 2015). In addition, we also include the population size itself as effects might differ by size of the origin country. Second, we construct a dummy variable which indicates if the country is a member of the European Union in a given year to control for immigration regulations. Third, we include data on bilateral trade flows, [5] i.e. exports from and imports to Germany, provided by Destatis (2018b) as a control for possible economic relations. Fourth, data on the migrant stock of the origin country in Germany from the 'Ausländerzentralregister' (Destatis 2018a) allows controlling for ethnic networks. Finally, to capture the effect of violent conflicts in the origin country on migration to Germany, we use the UCDP/PRIO Armed Conflict Dataset Version 4-2016 (Allansson et al. 2017; Gleditsch et al. 2002). We construct two dummy variables: the dummy variable "conflict" takes the value one in case of 25 to 999 battle-related deaths for a given origin country in a given year The dummy variable "war" takes the value one in case of more than 999 battle-related deaths for a given origin country in a given year.

3.4 Sample construction

To construct our dataset, we proceed as follows: First, we include all countries for which we have information on GDP, trade flows and population size for all years between 1977 and 2014. Second, we restrict the sample according to the availability of migration data: We only include countries for which we have migration data in all years but 1990, 2000 and 2001. In these three years, migration data of Destatis included many missing observations due to changes in the data generation process. We interpolate these missing observations linearly on the basis of the years 1989 and 1991, and 1999 and 2002, respectively. Third, we add the data from our GI dataset about the presence of the GI and the number of institutes per country and year, and assign the value 0 to these variables for countries and years that are not included in the GI dataset.

Furthermore, in some cases, the information about language services in the annual reports are reported jointly for two or more institutes without clarifying which of these institutes actually offered language services. This is mainly the case for institutes that are subsidiaries

[5] For an comprehensive overview of the interrelations between trade and migration see Felbermayr et al. (2015).



(a) Countries with GI (b) Institutes

Figure 2: Number of countries and institutes (based on 69 countries in the sample).

of other main institutes.[6] To avoid any ambiguity, we drop those countries where the number of institutes that offer language services is not clearly distinguishable from those that do not offer language services. As this problem mainly occurs before 1977, we limit the observation period to the years afterwards.

Finally, we end up with a balanced dataset that includes observations for 69 countries in the period from 1977 to 2014.[7]

3.5 Descriptive statistics

Table 1 provides summary statistics of the variables used in our analysis. Looking more closely at the presence of the GI at the country level in the period 1977 to 2014, we see that there was at least one institute in at least one year in 49 of the 69 countries; in 30 countries of them, the GI was present in all years of the observation period. The worldwide distribution of the countries in our sample is displayed in Figure A1 where countries are grouped according to the number of years in which the GI was present (all years of the observation period, at least one year and no year). The countries in our sample are spread over all continents and so are those with presence of the GI.[8]

Figure 2a presents the number of countries per year in which the GI was present with any type of institute or, respectively, with at least one institute that offered language services. In addition, the number of countries without any institute in that year is included. In most of the years, the GI was present in around 40 out of the 69 countries. Throughout the entire period, the number of countries with institutes which only offered non-language services was negligible.

A different picture emerges when we compare the number of institutes per year with and without language services (Figure 2b). While in the entire period 1977 to 2014, the number

[6]For more information on the different types of institutes in the dataset, please refer to Uebelmesser et al. (2018a).

[7]For a list with the names of the 69 countries, see Table A1 in the Appendix.

[8]Note that the (former) Soviet Union and other former socialist countries are not included in our sample. This is due to many newly founded states in the beginning of the 1990s and the lack of GDP and migration data. Furthermore, some Western European countries are missing as joint reporting has been a widespread phenomenon in these countries. As a robustness check, we use an unbalanced sample which allows considering many more countries including former socialist countries and we also present two specifications where we include countries with joint reporting (see Section 5.2).

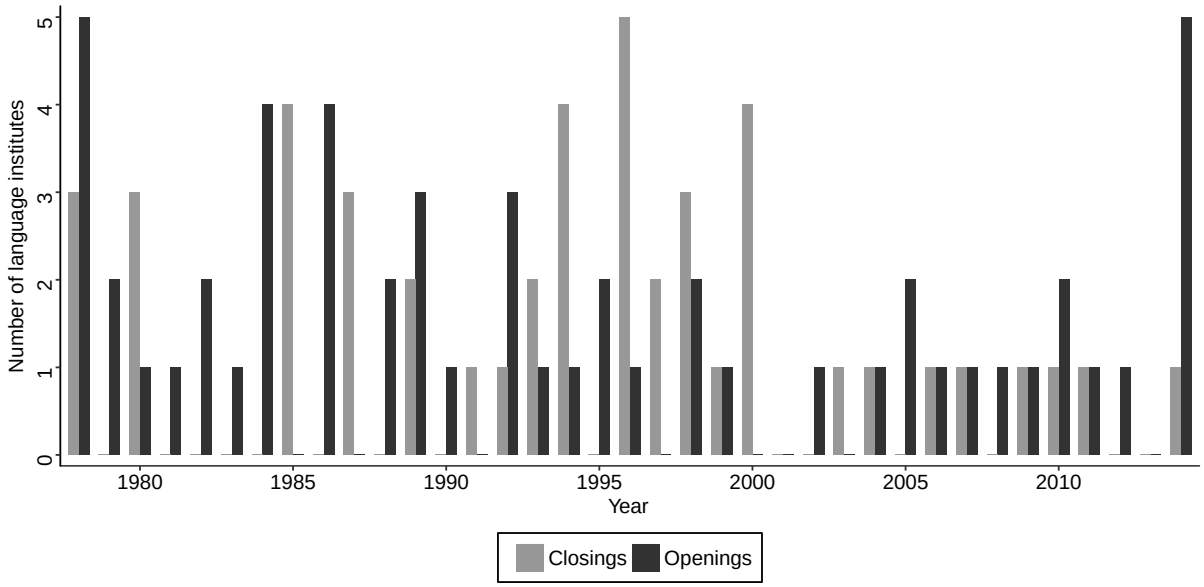


Figure 3: Number of openings and closings (based on 69 countries in the sample).

of institutes which offered language services always exceeded the number of institutes that did not, there was a significant number of institutes of the latter group.

The aggregate numbers hide that there is substantial variation in each year due to the closings and openings of institutes (see also Table 1 for the overall, between and within variation). Figure 3 shows this for institutes with language services.[9]

4 Theoretical background and empirical strategy

4.1 A random utility maximization model of migration

The micro-foundation of migration choice can be modelled in a RUM model. We describe a simple RUM model based on Bertoli and Fernández-Huertas Moraga (2013). Individual i in origin country j decides to locate in country k out of the set of alternative destination countries D which maximizes his utility

$$U_{ijk} = V_{jk} + \epsilon_{ijk} = w_{jk} - c_{jk} + \epsilon_{ijk} \tag{1}$$

where V_{jk} is the deterministic part of the utility capturing expected income, w_{jk}, which increases individual utility, and costs of migration, c_{jk}, such as geographical distance or strict visa regulations, which decrease utility. ϵ_{ijk} is the error term, which will be explained in more detail below. Language learning likely increases expected earnings in the destination country by improving migrants' labour-market outcomes, but also raises the costs of migration which include the costs of language acquisition. Individuals will opt for language learning if their utility is maximized by doing so. Following Bertoli and Fernández-Huertas Moraga (2013), these individual decisions are not explicitly included in the model, but they are captured in the error term. Given that the unit of our analysis of

[9]Closing does not necessarily mean that the whole institute was closed, but that language services were no longer offered. The same holds for opening.

the determinants of migration will be the country level, we include measures for foreign language learning in the deterministic part of the utility V_{jk} aggregated at the country level.

The resulting choice probabilities for one of the alternatives are determined by the assumptions on the distribution of the error term ϵ_{ijk}. Grogger and Hanson (2011), Mayda (2010) and others adopted the standard multinomial logit model, assuming that ϵ_{ijk} follows an independent and identically distributed extreme value type 1 distribution (McFadden 1974). Then, the probability of migrating from country j to k is

$$p_{ijk} = \frac{e^{V_{jk}}}{\sum_{l \in D} e^{V_{jl}}}. \tag{2}$$

The assumption on the distribution of the error term implies that the denominator $\sum_{l \in D} e^{V_{jl}}$ does not vary across origin countries. In other words, the attractiveness of destination country k is the same for all individuals across all origin countries j. The ratio of the probabilities of migrating to country k and of not migrating (i.e. staying in the country of origin j) shows this property, known as the Independence of Irrelevant Alternatives (IIA) assumption,

$$\frac{p_{ijk}}{p_{ijj}} = \frac{e^{V_{jk}}}{e^{V_{jj}}}. \tag{3}$$

This ratio depends only on the relative attractiveness of countries j and k, but not on any alternative destination. In our set-up where the focus is on the presence of German language learning opportunities in countries of origin on migration to Germany, the IIA assumption can be regarded to be less restrictive.[10]

4.2 Estimation method

For the reasons outlined above, we build on the log-linearization of Equation (3) for our basic empirical specifications. We can ignore the empirical counterpart of V_{jk} as Germany is the only destination country in our estimation. We therefore estimate the following fixed-effects model

$$y_{jt} = \alpha' GI_{jt} + \boldsymbol{\beta'} \boldsymbol{x_{jt}} + \boldsymbol{\phi_t'} \boldsymbol{d_t} + \boldsymbol{\phi_j'} \boldsymbol{d_j} + \boldsymbol{\phi_j'} \boldsymbol{d_{jT}} + \eta_{jt} \tag{4}$$

where y_{jt} represents the logarithm of the migration flow from origin country j to Germany in year t over the number of people that stay in origin country j. GI represents our main variable of interest, the number of (language) institutes. $\boldsymbol{x}_{jt}$ is the vector of control variables, that includes the log of the population in origin country j, the log of the GDP/capita as a measure for economic conditions, variables that indicate conflicts and

[10] In Section 6.3, we relax this assumption and control for multilateral resistance to migration. This concept was introduced by Bertoli and Fernández-Huertas Moraga (2013) on the basis of the generalized nested logit model (Wen and Koppelman 2001) and allows for correlation of the error terms across alternative destinations and thus for changes in alternative destinations, which are substitutes, to affect bilateral migration flows.

wars, the log of the existing migrant stock of country j in Germany in $t-1$ as a measure for previous migration to Germany, a dummy that indicates the EU membership and the log of the trade volume (import and exports) with Germany. Furthermore, we include a set of dummies to control for fixed effects. We add year dummies d_t to control for origin-invariant effects, origin dummies d_j to control for all time-invariant characteristics of the origin country and relations between Germany and the origin country j. We further add dummies d_{jT} that vary by origin country j and ten-year-time periods T. As our time frame covers 38 years, these fixed effects help to control for slowly changing factors in the relations between Germany and the origin country. η_{jt} are the error terms. Standard errors are clustered on the country level. We weight observations by the population of the origin countries to ensure that each potential migrant receives the same weight in the estimations, independent of the origin.

5 Results for migration flows to Germany

We estimate Equation (4) in several specifications. Our basic specifications in Table 2 employ the number of language institutes, i.e. institutes that offer language services. In Table 3, robustness checks are presented including lagged specifications. Table 4 shows that decisions regarding the sample choice do not influence the results. In Table 5, we study whether effects differ for countries with specific characteristics like geographic, linguistic and economic distance, as well as for countries with conflicts and wars. Finally, Figure 4 shows evidence how the effect has evolved over time. As we include origin-country fixed effects and origin-country*10-year fixed effects in all specifications, the estimated effects capture within-country-10-year variations.

5.1 Basic specifications

For the basic specifications presented in Table 2, we include the control variables in a stepwise way. For all specifications, the association between the number of language institutes and the migration rate to Germany is positive and highly significant. One more language institute is associated with an average increase of the migration rate to Germany by $e^{0.124} - 1 \approx 13.2\%$.[11]

The coefficients of log GDP/capita are negative and significant in all specifications. Better economic conditions in the origin country are thus negatively related with the benefits of migration. Furthermore, EU-membership is significantly and positively correlated with migration to Germany, which might capture less restrictive immigration regulations due to the principle of free movement of labour which holds within the EU and which lowers migration costs. The log of population is not significantly correlated with the migration rate in any specifications, neither are war, conflict or trade flows. Finally, the migrant stock of the origin country already present in Germany is positively and highly significantly associated with the migration rate to Germany. This might capture network effects, as the lagged migration rate is a measure for past migration from the origin country to Germany.

[11] The results are robust if observations are not weighted by population size of the origin country (see Table A2 in the Appendix).

Table 2: Estimation results: basic specifications

DV: log migration rate	(1)	(2)	(3)	(4)
Language institutes	0.107*	0.111*	0.112**	0.124**
	(0.0598)	(0.0570)	(0.0548)	(0.0538)
log GDP per capita		-0.368**	-0.435***	-0.356**
		(0.183)	(0.160)	(0.147)
EU member		0.762***	0.756***	0.652***
		(0.243)	(0.239)	(0.176)
log population		-0.579	-0.554	-0.279
		(0.937)	(0.939)	(0.695)
Conflict			-0.0302	-0.0218
			(0.0786)	(0.0726)
War			0.0524	0.0725
			(0.0928)	(0.0888)
log (exports + imports)			0.0900	0.0770
			(0.0835)	(0.103)
log (migrant stock / population), lag =1				0.413***
				(0.145)
Constant	-11.66	2.097	0.299	-1.305
	(178,843)	(194,647)	(193,030)	(125,296)
Observations	2,622	2,622	2,622	2,622
Number of country	69	69	69	69
Years	1977-2014	1977-2014	1977-2014	1977-2014
Adjusted R-squared	0.611	0.623	0.626	0.658
Year FE	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes
Country*10-year FE	Yes	Yes	Yes	Yes

Cluster-robust standard errors in parentheses. Observations are weighted by population size.
*** p<0.01, ** p<0.05, * p<0.1.

5.2 Robustness Checks

5.2.1 Independent variable choice

In Table 3, we present robustness checks based on our preferred specification (Column (4) in Table 2). First, the positive correlation between the number of language institutes and the migration rate might not measure the effect of language learning opportunities but other aspects that come with the opening of a new institute by the GI. Beyond language services, the institutes provide information about German culture and society. This might reduce uncertainty about life in Germany for potential migrants and therefore increase migration to Germany. The first two columns in Table 3 take this into account. In Column (1), we replace the number of language institutes in a given country and year with the number of all institutes, with and without language services. We find that the coefficient on the migration rate is reduced to 0.0645. In Column (2), we split the total number of institutes into two independent variables – the number of institutes with language services and the number of institutes without language services. We find that only additional language institutes significantly affect the migration rate, whereas institutes without language services do not. The coefficient for the number of language institutes

is the same as in our preferred specification. From the results of these two robustness checks, we conclude that our variable of interest, when it comes to the empirical relation with migration to Germany, is indeed the variable which measures language learning opportunities and not other effects of the GI.

Furthermore, we include additionally the first lag of the number of language institutes (Column (3)), and the first and second lags (Column (4)). Language learning and migration might not take place in the same period, as the acquisition of language skills requires some time. Therefore, the effect of an additional institute might be better captured with lags as it might evolve over time. This is indeed what we find: a new institute has an even larger effect on the migration rate to Germany after one year, as we can see in Column (3), while for higher lags there are no further significant effects (see Column (4)). In Column (5), we lag all control variables by one year, as individuals might react to conditions in the previous year. The effect of language institutes does not change compared to our preferred specification.

Additionally, it might be the case that only large enough institutes are actually able to influence the migration rate and therefore drive our results. Unfortunately, we cannot measure the actual size of institutes in terms of course participation, since there is no consistent measure available over this long period.[12] In Column (6), we therefore relate the number of institutes in a country to the population size. This controls for the possibility that the effect on the migration rate of one more institute in, for example, India and one more institute in Iceland might be different. We find that the coefficient for the log of the number of language institutes per 1m inhabitants remains positive and significant.

5.2.2 Sample choice

Table 4 shows that the results do not depend on choices taken regarding the sample. In Columns (1) and (2), we also include countries for which it is not always entirely clear for some institutes if they have offered language services or not, because of joint reporting in the annual reports. In those cases, annual reports show numbers on language course and exam participation jointly for a first-named institute (mostly main institute) followed by one or more institutes in parenthesis which are mainly subsidiaries. The independent variable in Column (1) assumes, that only the first-named institutes offered language services. In Column (2), it is assumed, that all institutes offered language services, first-named institutes and not-first-named institutes. In these cases, the sample increases to 86 countries (see map in Figure A2). In both columns the effect is slightly smaller, but does not change qualitatively.

Column (3) shows an unbalanced sample including all time periods for all country with at least ten consecutive observations. This enlarges the sample in two ways. First, we no longer drop all observations of a country if in some years joint reporting occurred. Second, additional countries are added to the sample, which came into existence later than 1977, like successors of the Soviet Union or Yugoslavia. This sample then contains 153 countries (see map in Figure A3). Again, we find that this does not change our results qualitatively.

[12]For data about language course participation see Uebelmesser et al. (2018a)

Table 3: Robustness checks

DV: log migration rate	(1)	(2)	(3)	(4)	(5)	(6)
All institutes	0.0645**					
	(0.0277)					
Institutes without lang. serv.		0.00435				
		(0.0246)				
Language institutes		0.124**	0.0824**	0.0638***		
		(0.0547)	(0.0381)	(0.0189)		
Language institutes, lag=1			0.0891**	0.111**	0.125**	
			(0.0405)	(0.0536)	(0.0520)	
Language institutes, lag=2				-0.00551		
				(0.0255)		
log (lang. institutes/pop)						0.0557**
						(0.0251)
Other controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,622	2,622	2,553	2,484	2,553	2,622
Number of country	69	69	69	69	69	69
Years	1977-2014	1977-2014	1977-2014	1977-2014	1977-2014	1977-2014
Adjusted R-squared	0.654	0.658	0.653	0.662	0.646	0.654
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes	Yes	Yes
Country*10-year FE	Yes	Yes	Yes	Yes	Yes	Yes
Lagged controls	No	No	No	No	Yes	No

Cluster-robust standard errors in parentheses. Observations are weighted by population size. Other controls (log (GDP/capita) EU member, log population, conflict, war, log(exports+imports), first lag of log(migrant stock/population)) and constant are included, but not shown. *** p<0.01, ** p<0.05, * p<0.1.

Table 4: Robustness checks: joint reporting of institutes and unbalanced sample

DV: log migration rate	(1)	(2)	(3)
Language institutes	0.105**	0.0967**	0.105*
	(0.0419)	(0.0416)	(0.0544)
Other controls	Yes	Yes	Yes
Observations	3,268	3,268	4,412
Number of country	86	86	153
Years	1977-2014	1977-2014	1977-2014
Adjusted R-squared	0.649	0.649	0.639
Year FE	Yes	Yes	Yes
Country FE	Yes	Yes	Yes
Country*10-year FE	Yes	Yes	Yes

Cluster-robust standard errors in parentheses. Observations are weighted by population size. Other controls (log GDP per capita, EU member, log population, conflict, war, log(exports+imports) and first lag of log(migrant stock/population)) and constant are included, but not shown.*** p<0.01, ** p<0.05, * p<0.1.

5.2.3 Heterogenous effects

As our a sample covers a long time period and a large number of countries with distinct characteristics, we test if there are heterogenous effects of the number of language institutes on migration to Germany for different groups of countries, and how these effects change over time.

First, our results might be driven by different country-groups due to different time-invariant characteristics, which are absorbed by our fixed effects. To better understand whether this is the case, we estimate our preferred specification (Column (4) of Table 2) with an additional interaction term for the GI variable and a dummy variable that captures geographic and linguistic distance, respectively. The dummy for geographical distance (Mayer and Zignago 2011) takes the value one if the distance to Germany is larger than or equal to the median distance to Germany. In an analogous way, linguistic distance is captured by a dummy variable which takes the value one if the major language spoken in the country is a non-Germanic language which holds for 60 countries in our sample (Adserà and Pytliková 2015).

We also explicitly consider the possible interactions between the GI variable and economic distance. We construct a dummy that indicates for each year if the log GDP/capita is smaller than or equal to the median log GDP/capita. Furthermore, interactions between the GI variable and dummy variables for the presence of conflicts and wars are studied.

Table 5 presents the results, where Column (1) shows again the preferred specification for better comparison. We can see that economic distance matters, i.e. the effect of the number of institutes is significantly higher for countries with lower income, even though the association is still positive and significant for both groups of countries. The same holds for linguistic distance. On the contrary, geographic distance does not seem to affect the results.

Finally, conflicts and wars have a different impact on the effect. While there is no heterogeneity in the association between the number of language institutes and the migration rate for countries with conflicts, we find that for countries with wars: the relationship between the number of institutes and migration to Germany is smaller on the 10 % significance level for countries where wars take place.

While conflicts might not be strong enough to change the effect, for countries with wars this might capture cases where due to wars, the institutes, and more generally language learning, plays a smaller role for migration compared to countries without a war. It might also capture a situation where due to wars institutes had to close because of violent conflicts, while at the same time these wars increased migration flows to Germany.

Second, the effect might change over time. Therefore, we estimate our basic specification for sub-samples of 20 years. Figure 4 plots the coefficients for the number of language institutes and the 95 % and 90 % confidence intervals for different 20-year sub-samples where the indicated year marks the first year of the respective 20-year period. While the effect is quite large in the sub-sample of 1977 (0.188), this effect decreases in the subsequent sub-samples until the sub-sample of 1981 (0.074). The effects in the following sub-samples stay roughly constant until the sub-samples starting at the beginning of the 1990s. Then, there is a further slight decrease until the last sub-sample (0.037). While the

Table 5: Heterogenous effects

DV: log migration rate	(1)	(2) Economic distance	(3) Geographic distance	(4) Linguistic distance	(5) Conflict	(6) War
Language institutes	0.124**	0.0705***	0.0893**	0.0582**	0.127**	0.126**
	(0.0536)	(0.0261)	(0.0401)	(0.0227)	(0.0545)	(0.0569)
Language institutes*...		0.263***	0.148	0.213***	0.0167	-0.0404*
		(0.0565)	(0.0919)	(0.0569)	(0.0163)	(0.0213)
Other controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,622	2,622	2,622	2,622	2,622	2,622
Countries	69	69	69	69	69	69
Years	1977-2014	1977-2014	1977-2014	1977-2014	1977-2014	1977-2014
Adjusted R-squared	0.661	0.669	0.663	0.666	0.662	0.664
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Country*10year FE	Yes	Yes	Yes	Yes	Yes	Yes

Cluster-robust standard errors in parentheses. Observations are weighted by population size. Other controls (log GDP per capita, EU member, log population, conflict, war, log(exports+imports) and first lag of log log(migrant stock/population)) and constant are included, but not shown. *** p<0.01, ** p<0.05, * p<0.1.

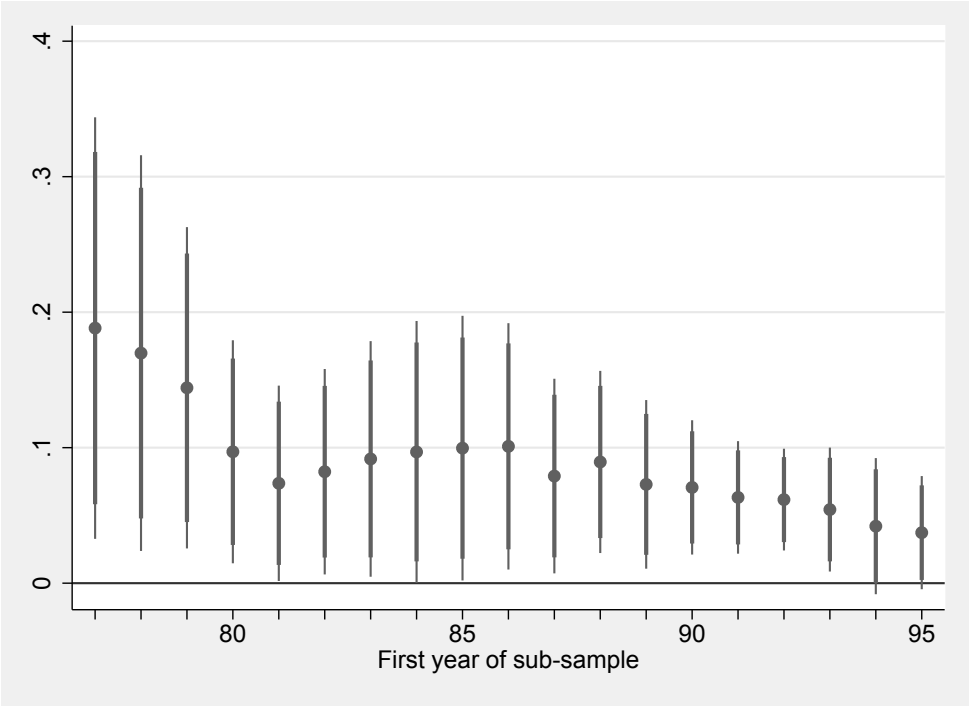


Figure 4: Coefficient plot for the number of language institutes: 20-years sub-samples (95 % and 90 % confidence intervals)

coefficients are positive and significant at least on the 5 % level in most sub-samples, they are only significant at the 10 % level in the last two sub-samples. This points towards a smaller role of language learning opportunities provided by the GI for migration which can be related to a more global world and, among others, to the advent of online language learning platforms in the late 2000s.

6 Threats to exogeneity

In the previous section, we have shown that there is a robust positive relationship between language learning opportunities and migration flows to Germany. In order to interpret that causally we have to deal with three potential threats to exogeneity: reverse causality, omitted variable bias and multilateral resistance to migration.

6.1 Reverse Causality: The Case of Switzerland

So far, we have found a positive relationship between the number of language institutes and the migration rate. However, we cannot derive from this result the direction of causality. Even though there is evidence that the location decision is based on a large number of considerations unrelated to migration (see Section 2), we cannot exclude that the GI is more likely to open institutes in countries with larger migration to Germany. It might then be impossible to disentangle the positive correlation of the GI and the migration rate, as estimated in Section 5, in the "migration effect" caused by the opening of an institute on the migration rate and the "selection effect" caused by the migration rate on the location decision for a new GI. As a consequence, we would expect the relationship between the number of institutes and the migration rate to be overestimated, i.e. the migration effect would be biased upwards.

We deal with this issue by estimating migration flows to Switzerland. The GI and the FFO neither consult with the Swiss government, nor are they influenced by Swiss representatives, as far as the decision process for opening and closing institutes is concerned.[13] The number of institutes can therefore be considered exogenous to migration flows to Switzerland.

In more detail, we replace the dependent variable used so far by the migration rate to Switzerland. We distinguish between the German-speaking part and the non-German-speaking part of Switzerland.[14] By focusing on Switzerland, we can exclude reverse causality as language institutes are exogenous to migration flows to Switzerland and we can see if language is the driver behind the observed effects. We expect that language institutes have a positive and significant effect on migration flows to the German-speaking part, but neither to the Italian- nor the French-speaking part. The effect on migration to the non-German-speaking part can also be seen as a placebo-test for the treatment of language institutes[15].

[13]The GI cooperates with the Swiss and Austrian embassies in some host countries once institutes have been opened

[14]We define a canton as German-speaking if the main language for the majority was German in 2010 according to the Bundesamt für Statistik (2012) and then aggregate the data for German- and non-German-speaking cantons.

[15]Eugster et al. (2017) show that the French- and Italian-speaking part of Switzerland is indeed culturally different to the German-speaking part. Therefore, it can be seen as a "different" country and can be used in a placebo-test.

Table 6: Estimation results - Switzerland

DV: log migration rate	(1) DEU	(2) CHE (German)	(3) CHE (non-German)	(4) CHE (German)	(5) CHE (non-German)
Language institutes	0.0352**	0.106**	0.0622***	0.205**	0.00681
	(0.0169)	(0.0403)	(0.0179)	(0.0862)	(0.0393)
Other controls	Yes	Yes	Yes	Yes	Yes
Observations	1,587	1,587	1,587	1,311	1,311
Countries	69	69	69	57	57
Years	1992-2014	1992-2014	1992-2014	1992-2014	1992-2014
Adjusted R-squared	0.702	0.668	0.664	0.570	0.584
Year FE	Yes	Yes	Yes	Yes	Yes
Country*10-year FE	Yes	Yes	Yes	Yes	Yes

*** p<0.01, ** p<0.05, * p<0.1; Robust standard errors in parentheses observations are weighted by population size of the origin. Other controls (log(GDP/capita), EU/EFTA member, log population, conflict, war, log(exports+ imports) and first lag of log(migrant stock/population)) and constant are included, but not shown. In (4) and (5), countries are excluded with significantly (at least on the 10 %-level) related migration flows to Germany and to the German-speaking part of Switzerland that have a variation in the number of language institutes in the period 1992 – 2014. Additionally, South Korea is excluded.

We estimate Equation 4 with data for Switzerland. Migration flows to Switzerland are provided by the Bundesamt für Statistik (2016) and Bundesamt für Statistik (2017). We use the same sample of countries as for Germany as destination country, but only for the period 1992 – 2014 due to data availability. To control for origin-specific characteristics that are not bilaterally related to Switzerland, we use the same variables as for the German specification, i.e. GDP/capita, population, conflict and war. Furthermore, we construct a variable for the free movement of persons between the EU/EFTA (Liechtenstein, Iceland and Norway) and Switzerland (Staatssekretariat für Migration 2017). The variable takes the value one if there is some relaxation of the immigration rules and zero otherwise. These relaxations include different steps towards free movement, like quotas, national worker priority or an invocation of the safeguard clause. Data on the migrant stocks is based on migration flow data. Exports and imports to Switzerland are also included, but cannot be disaggregated for the German-speaking and non-German-speaking part of Switzerland (Barbieri and Keshk 2016; Barbieri et al. 2009).

Table 6 shows the results. In Column (1), we re-run our preferred specification for Germany, albeit for the shorter period from 1992 - 2014, in order to allow comparing the results for Germany and Switzerland. Compared to our previous estimations with the longer period, we find that the effect size of the coefficient for the number of language institutes is reduced to 0.035 and is significant on the 5 % level. This results is in line to what we can see in Figure 4 for the corresponding sub-samples.

Column (2) presents the results for the German-speaking part of Switzerland and Column (3) presents the results for the non-German-speaking part of Switzerland. In both cases the language institutes have a positive and significant effect on migration flows, however the size of the coefficient is different; while for the German-speaking part the coefficient is 0.106, the coefficient for the non-German-speaking part is substantially lower with 0.062.

However, we have to address one further concern: migration flows to Germany and to Switzerland could be correlated. The decision to open or close a language institute by FFO and GI would then no longer be exogenous to migration to Switzerland.

Therefore, we re-run the estimations for Switzerland but exclude countries from the sample with significantly (at least on the 10 %-level) positively or negatively related migration flows to Germany and to the German-speaking part of Switzerland with a variation in the number of language institutes in the period 1992 – 2014.[16] Furthermore, we drop Korea, which constitutes an outlier that strongly biases the results.[17]

Column (4) shows that the size of the effect of language institutes on the migration rate to the German-speaking part is almost doubled for this reduced sample. The significant correlation between migration flows to Germany and to the German-speaking part of Switzerland, which is present for some countries, does not affect our main results in an important way. In Column (5), the same sample as in Column (4) is estimated with migration flows to the non-German speaking part of Switzerland as dependent variable. In comparison to Column (3), the effect decreases almost to zero and loses significance.

As we exclude any direct effect of migration flows to Switzerland on the opening and closing decision for institutes, and as there is also no evidence of an indirect effect via positively correlated migration flows to the German-speaking part of Switzerland and to Germany, we interpret the result for the German-speaking part of Switzerland (see Column (4)) as a causal effect of the presence of language learning opportunities abroad on migration from those countries to Switzerland.

6.2 Omitted variable bias: student mobility, trade relations and children-age language learning

There might be an omitted variable bias if we do not control for factors that drive both, the number of institutes and the migration flows to Germany. In our case, this is only problematic if these factors change within ten-year periods. Otherwise, we are able to absorb the potential endogeneity by the included fixed effects that vary over countries and ten-year periods. To test for a potential omitted variable bias, we include additional control variables.

First, we focus on student migration via the presence of the German Academic Exchange Service (Deutscher Akademischer Austauschdienst, DAAD), which supports student mobility and academic exchange with Germany. We construct a dummy variable based on the presence of regional DAAD offices and DAAD information centres.

Second, we control for established trade relationships and networks by including a dummy variable for the presence of German Chambers of Commerce Abroad (Deutsche Auslandshandelskammern, AHK). The AHK carry out foreign trade promotion and support

[16] This is based on country-specific regressions of the log of the migration rate to Switzerland on the log of the migration rate to Germany. The following 11 countries are excluded: Costa Rica, Denmark, Finland, India, Iraq, Mexico, Norway, Poland, Sauda Arabia, USA, Vietnam.

[17] In 2014, four language institutes were opened in Korea. If we include Korea, the coefficient for the estimation of Column (4) in Table 6 is only 0.104 and insignificant. If we have a look at the corresponding 20-year sub-samples, we find that only in the last sub-sample from 1995 to 2014 the effect is insignificant, while the other sub-samples show similar results as Column (4) in Table 6.

Table 7: Estimation results - omitted variable bias

DV: log migration rate	(1)	(2)	(3)	(4)
Language institutes	0.123**	0.124**	0.125**	0.124**
	(0.0552)	(0.0534)	(0.0561)	(0.0568)
DAAD	0.0440			0.0444
	(0.147)			(0.149)
AHK		0.0424		0.0405
		(0.0996)		(0.106)
DAS			0.0103	0.00973
			(0.0601)	(0.0637)
Other Controls	Yes	Yes	Yes	Yes
Observations	2,622	2,622	2,622	2,622
Countries	69	69	69	69
Years	1977-2014	1977-2014	1977-2014	1977-2014
Adjusted R-squared	0.658	0.658	0.658	0.657
Year FE	Yes	Yes	Yes	Yes
Country*10-year FE	Yes	Yes	Yes	Yes

Cluster-robust standard errors in parentheses. Observations are weighted by population size. Other controls (log GDP per capita, EU member, log population, conflict, war, log (exports+imports), first lag of log(migrant stock/population)) and constant are included, but not shown. *** p<0.01, ** p<0.05, * p<0.1.

German companies in doing business abroad.

Third, we consider other ways of German language learning abroad, in particular language learning at school. However, there is a lack of comprehensive data.[18] As a proxy, we control for the number of German schools abroad (Deutsche Auslandsschulen, DAS) present in the country.

All three variables are significantly and positively correlated with the presence of the Goethe-Institut in a country. In Table 7 we check if including these variables affects the coefficient of the number of language institutes in our baseline specification. In Column (1), the variable for DAAD is included, in Column (2) the variable for AHK and in Column (3) the variable for DAS. Neither of them has a significant effect on the migration rate and more importantly, the effect of language institutes remains unchanged. This also holds for a specification (Column 4), where we jointly include all three additional variables. Hence, we conclude, that our results are robust to omitted variable bias.

6.3 Multilateral resistance to migration

As described in Section 4.1, multilateral resistance of migration might be another threat to exogeneity. Bertoli and Fernández-Huertas Moraga (2013) show that the CCE estimator

[18]The only relatively comprehensive data sources on German language learning are Ständige Arbeitsgruppe Deutsch als Fremdsprache (2003), Ständige Arbeitsgruppe Deutsch als Fremdsprache (2006), Netzwerk Deutsch (2010), and Auswärtiges Amt (2015) These data, however, are not suitable for a quantitative analysis as they are only available in 5-year steps since 1995, rather incomplete for the set of countries we have in our sample and often rely on estimation.

Table 8: Estimation results - common correlated effects

DV: log migration rate	(1) CCE	(2) CCE	(3) CCE	(4) CCE
Number of language institutes	0.0737***	0.0665***	0.0704***	0.0614***
	(0.0190)	(0.0177)	(0.0166)	(0.0174)
log GDP per capita		-0.484***	-0.515***	-0.563***
		(0.0474)	(0.0603)	(0.0610)
EU member		0.693***	0.896***	0.877***
		(0.223)	(0.215)	(0.211)
log population		-2.029***	-2.334**	-3.582***
		(0.747)	(0.927)	(1.053)
Conflict			-0.0172	-0.00759
			(0.0222)	(0.0228)
War			0.0180	0.0404
			(0.0329)	(0.0337)
log (exports + imports)			0.119***	0.127***
			(0.0273)	(0.0285)
log (migrant stock / population), lag =1				-0.0466*
				(0.0280)
Constant	-14.13*	-17.33	-3.581	19.40
	(7.796)	(54.36)	(58.42)	(88.92)
Observations	2,622	2,622	2,622	2,622
Number of country	69	69	69	69
Years	1977-2014	1977-2014	1977-2014	1977-2014
Adjusted R-squared	0.962	0.975	0.981	0.981
Year FE	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes
Country*10-year FE	Yes	Yes	Yes	Yes
CCE-Test (p-value)	0.00	0.00	0.00	0.00

Observations are weighted by population size; results are estimated with the CCE-estimator (Pesaran 2006).
*** p<0.01, ** p<0.05, * p<0.1.

by Pesaran (2006) consistently corrects for multilateral resistance to migration. Even though we think, that in our case this issue is of minor importance, we add a specification where we control for multilateral resistance of migration.

For the CCE estimator, cross-sectional averages of all dependent and independent variables interacted with heterogeneous coefficients for all countries have to be included. For our multiple-origin and single-destination setting we estimate the following equation

$$y_{jt} = \alpha' GI_{jt} + \boldsymbol{\beta'} \boldsymbol{x_{jt}} + \boldsymbol{\phi_t'} \boldsymbol{d_t} + \boldsymbol{\phi_j'} \boldsymbol{d_j} + \boldsymbol{\phi_j'} \boldsymbol{d_{jT}} + \boldsymbol{\lambda_j'} \tilde{\boldsymbol{z}}_{\boldsymbol{t}} + \eta_{jt} \tag{5}$$

with the weighted cross-sectional average defined as

$$\tilde{\boldsymbol{z}}_{\boldsymbol{t}} = \frac{1}{\sum_j \omega_{jt}} \left(\sum_j \omega_{jt} y_{jt}, \sum_j \omega_{jt} x_{jt} \right)$$

where ω_{jt} gives the weight for country j in t, for which we use population size.

In Table 8, we present results of our basic specifications (Table 2), but estimated with the CCE estimator according to Equation 5. The coefficient of the number of language institutes becomes smaller, but remains highly significantly different from zero.[19]

7 Conclusion

In this paper, we have analyzed how the presence of German language learning opportunities abroad affects migration to Germany. We find a significant and positive correlation between the number of language institutes of the GI and migration rates to Germany and can show that the language learning opportunities are indeed the relevant channel. This relationship is stronger for countries with lower income and where the majority of the people does not speak a Germanic language. Wars reduce the size of the effect.

We further show that the effect can also be found for migration to the German-speaking part of Switzerland. This excludes the possibility of reverse causality and thus allows interpreting the relationship as causal. While migration to Germany might be endogenous to the opening and closing of institutes, migration to Switzerland is not.

So far, similar effects have been only shown for foreign language learning at school for EU countries (Aparicio Fenoll and Kuehn 2016). Compulsory language learning at schools, and, even more, linguistic properties are however hardly within the reach of policy-makers in the destination country. This is different for language learning opportunities for adults. In particular, cultural institutes like the GI are often closely linked to policy-makers of the destination countries. We have shown that migration can be affected by the supply of language learning opportunities for adults in terms of the number of migrants – albeit with a smaller effect for more recent years. Hence, we provide support for the plans by the German government – also in the context of the soon-to-be adopted new immigration law – to more actively target foreign skilled labour in order to cope with skill shortage due to demographic change. The importance of language proficiency and the special role played by the GI in this process are explicitly mentioned (Goethe-Institut 2018; Bundesregierung 2018). The effect on the composition of migrants and their actual language proficiency and its more general effects on the destination country is left for future research.

When interpreting the results and their magnitude, a few points should be taken into account: First, the migration data we use does not allow identifying different groups of migrants, e.g. students, retirees, or those migrating for work reasons or to join their family. Similarly, we cannot tell whether migration is permanent or temporary and whether non-permanent migration is circular or not. These are all important aspects if one wants to assess the overall consequences of migration and even more though those for the origin countries. Furthermore, we cannot determine the exact reasons for the decreasing effect in more recent years. In order to shed more light on this, micro-level data is needed. For this, surveys in selected institutes are planned as next step.

[19]The CCE estimator requires a balanced panel with at least 20 years and 30 countries for a consistent estimation, which limits its applicability in sub-sample analyses.

References

Adserà, Alícia and Mariola Pytliková (2015). "The Role of Language in Shaping International Migration". *The Economic Journal* 125(586), F49–F81.

Allansson, Marie, Erik Melander, and Lotta Themnér (2017). "Organized violence, 1989–2016". *Journal of Peace Research* 54(4), 574–587.

Aparicio Fenoll, Ainhoa and Zoë Kuehn (2016). "Does Foreign Language Proficiency Foster Migration of Young Individuals within the European Union?" In: *The Economics of Language Policy*. Ed. by Bengt-Arne Wickstroem and Michele Gazzola. CESifo Seminar Series. Cambridge, MA: MIT Press, 331–355.

Auswärtiges Amt (2015). *Deutsch als Fremdsprache weltweit. Datenerhebung 2015*. Berlin.

Barbieri, Katherine and Omar M.G. Keshk (2016). *Correlates of War Project Trade Data Set Codebook, Version 4.0*. URL: http://correlatesofwar.org.

Barbieri, Katherine, Omar M.G. Keshk, and Brian M. Pollins (2009). "Trading data: Evaluating our assumptions and coding rules". *Conflict Management and Peace Science* 26(5), 471–491.

Belot, Michèle and Timothy J. Hatton (2012). "Immigrant Selection in the OECD". *The Scandinavian Journal of Economics* 114(4), 1105–1128.

Bertoli, Simone and Jesús Fernández-Huertas Moraga (2013). "Multilateral resistance to migration". *Journal of Development Economics* 102, 79–100.

Bleakley, Hoyt and Aimee Chin (2010). "Age at Arrival, English Proficiency, and Social Assimilation Among US Immigrants". *American Economic Journal: Applied Economics* 2(1), 165–192.

Bundesamt für Statistik (2012). *Ausgewählte Indikatoren im regionalen Vergleich, 2012*. URL: https://www.bfs.admin.ch/bfs/de/home/statistiken/regionalstatistik/regionale-portraets-kennzahlen/kantone.assetdetail.2420204.html.

— (2016). *Statistik der ausländischen Wohnbevölkerung (1991-2009)*. URL: https://www.bfs.admin.ch/bfs/de/home/statistiken/bevoelkerung/erhebungen/petra.assetdetail.6805.html.

— (2017). *Einwanderung der ständigen Wohnbevölkerung nach Kanton, Staatsangehörigkeit (Kategorie), Herkunftsstaat, Geschlecht und Altersklasse*. URL: https://www.bfs.admin.ch/bfs/de/home/statistiken/bevoelkerung/migration-integration/internationale-wanderung.assetdetail.3202566.html.

Bundesregierung (2018). *Eckpunkte zur Fachkräfteeinwanderung aus Drittstaaten*. Berlin.

Chiswick, Barry R. and Paul W. Miller (1995). "The Endogeneity between Language and Earnings: International Analyses". *Journal of Labor Economics* 13(2), 246–288.

Destatis (2016). "Qualitätsbericht Wanderungen". Wiesbaden.

Destatis (2018a). *Ausländer: Deutschland, Stichtag, Geschlecht/Altersjahre/ Familienstand, Ländergruppierungen/Staatsangehörigkeit - Tabelle 12521-0002*. URL: https://www-genesis.destatis.de/genesis/online.

— (2018b). *Aussenhandel*. URL: https://www.destatis.de/DE/ZahlenFakten/GesamtwirtschaftUmwelt/Aussenhandel/Aussenhandel.html.

Dustmann, Christian and Francesca Fabbri (2003). "Language proficiency and labour market performance of immigrants in the UK". *The Economic Journal* 113(489), 695–717.

Dustmann, Christian and Arthur van Soest (2001). "Language Fluency and Earnings: Estimation with Misclassified Language Indicators". *Review of Economics and Statistics* 83(4), 663–674.

Eugster, Beatrix, Rafael Lalive, Andreas Steinhauer, and Josef Zweimüller (2017). "Culture, Work Attitudes, and Job Search: Evidence from the Swiss Language Border". *Journal of the European Economic Association* 15(5), 1056–1100.

Feenstra, Robert C., Robert Inklaar, and Marcel P. Timmer (2015). "The Next Generation of the Penn World Table". *American Economic Review* 105(10), 3150–3182.

Felbermayr, Gabriel, Volker Grossmann, and Wilhelm Kohler (2015). "Migration, International Trade, and Capital Formation: Cause or Effect?" In: *Handbook of the Economics of International Migration*. Ed. by Barry R. Chiswick and Paul W. Miller. Vol. 1. North-Holland, 913–1025.

Gleditsch, Nils Petter, Peter Wallensteen, Mikael Eriksson, Margareta Sollenberg, and Havard Strand (2002). "Armed Conflict 1946-2001: A New Dataset". *Journal of Peace Research* 39(5), 615–637.

Goethe-Institut (2014). *Jahrbuch 2013/14*. Munich.

— (2018). *Pressemappe Jahrespressekonferenz 2018*. München.

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

Mayda, Anna Maria (2010). "International migration: a panel data analysis of the determinants of bilateral flows". *Journal of Population Economics* 23(4), 1249–1274.

Mayer, Thierry and Soledad Zignago (2011). "Notes on CEPII's distances measures: The GeoDist database". *CEPII Working Paper* 2011-25, CEPII, Paris.

McFadden, Daniel (1974). "Conditional logit analysis of qualitative choice behavior". In: *Frontiers in Econometrics*. Ed. by Paul Zarembka. New York: Academic Press, 105–142.

Netzwerk Deutsch (2010). *Statistische Erhebungen 2010: Die deutsche Sprache in der Welt*. Berlin.

Pesaran, M. Hashem (2006). "Estimation and Inference in Large Heterogeneous Panels with a Multifactor Error Structure". *Econometrica* 74(4), 967–1012.

Staatssekretariat für Migration (2017). *Personenfreizügigkeit Schweiz – EU/EFTA*. URL: https://www.sem.admin.ch/sem/de/home/themen/fza_schweiz-eu-efta.html.

Ständige Arbeitsgruppe Deutsch als Fremdsprache (2003). *Deutsch als Fremdsprache weltweit. Datenerhebung 2000*. Berlin.

— (2006). *Deutsch als Fremdsprache weltweit. Datenerhebung 2005*. Berlin.

Uebelmesser, Silke and Severin Weingarten (2017). "A macro-level analysis of adult-age language learning". *CESifo Working Paper* 6511, CESifo, Munich.

Uebelmesser, Silke, Matthias Huber, and Severin Weingarten (2018a). "The German Language Worldwide: a New Data Set on Language Learning". *CESifo Economic Studies* 64(1), 103–121.

— (2018b). "The German Language Worldwide: Data - Language Learning Opportunities Abroad". Harvard Dataverse, V1. DOI: 10.7910/DVN/VSQBM7.

Wen, Chieh-Hua and Frank S. Koppelman (2001). "The generalized nested logit model". *Transportation Research Part B: Methodological* 35(7), 627–641.

Appendix

Table A1: Countries in the sample

Burundi*	India	Peru
Benin*	Ireland	Philippines
Burkina Faso*	Iran	Poland
Bulgaria	Iraq	Paraguay*
Bolivia	Iceland*	Romania
Chile	Jordan	Rwanda*
Côte d'Ivoire	Kenya	Saudi Arabia
Cameroon	Cambodia*	Senegal
Congo - Brazzaville	South Korea	Sierra Leone*
Colombia	Liberia*	El Salvador*
Costa Rica	Sri Lanka	Sweden
Denmark	Madagascar*	Chad*
Dominican Republic*	Mexico	Togo
Algeria	Mali*	Thailand
Ecuador*	Malaysia	Trinidad & Tobago*
Ethiopia	Niger*	Tunisia
Finland	Nigeria	Turkey
Ghana	Nicaragua*	Tanzania
Guinea*	Norway	Uruguay
Guatemala*	Nepal	United States
Honduras*	New Zealand	Venezuela
Haiti*	Pakistan	Vietnam
Hungary	Panama*	South Africa

* Countries which never had a language institute between 1977 and 2014

Table A2: Basic specification unweighted

DV: log migration rate	(1)	(2)	(3)	(4)
Number of language institutes	0.128*	0.124*	0.127*	0.129**
	(0.0725)	(0.0677)	(0.0665)	(0.0618)
log GDP per capita		-0.534***	-0.580***	-0.410***
		(0.185)	(0.202)	(0.144)
EU member		0.927***	0.920***	0.774***
		(0.232)	(0.236)	(0.166)
log population		-1.620*	-1.628*	-0.768
		(0.843)	(0.843)	(0.641)
Conflict			0.0446	0.0501
			(0.0676)	(0.0591)
War			0.160*	0.159*
			(0.0880)	(0.0850)
log (exports + imports)			0.0577	0.0550
			(0.0582)	(0.0699)
log (migrant stock / population), lag =1				0.532***
				(0.127)
Constant	-10.66***	19.63	19.02	8.734
	(0.0768)	(14.28)	(14.28)	(10.56)
Observations	2,622	2,622	2,622	2,622
Number of country	69	69	69	69
Years	1977-2014	1977-2014	1977-2014	1977-2014
Adjusted R-squared	0.619	0.644	0.647	0.686
Year FE	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes
Country*10-year FE	Yes	Yes	Yes	Yes

Cluster-robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1

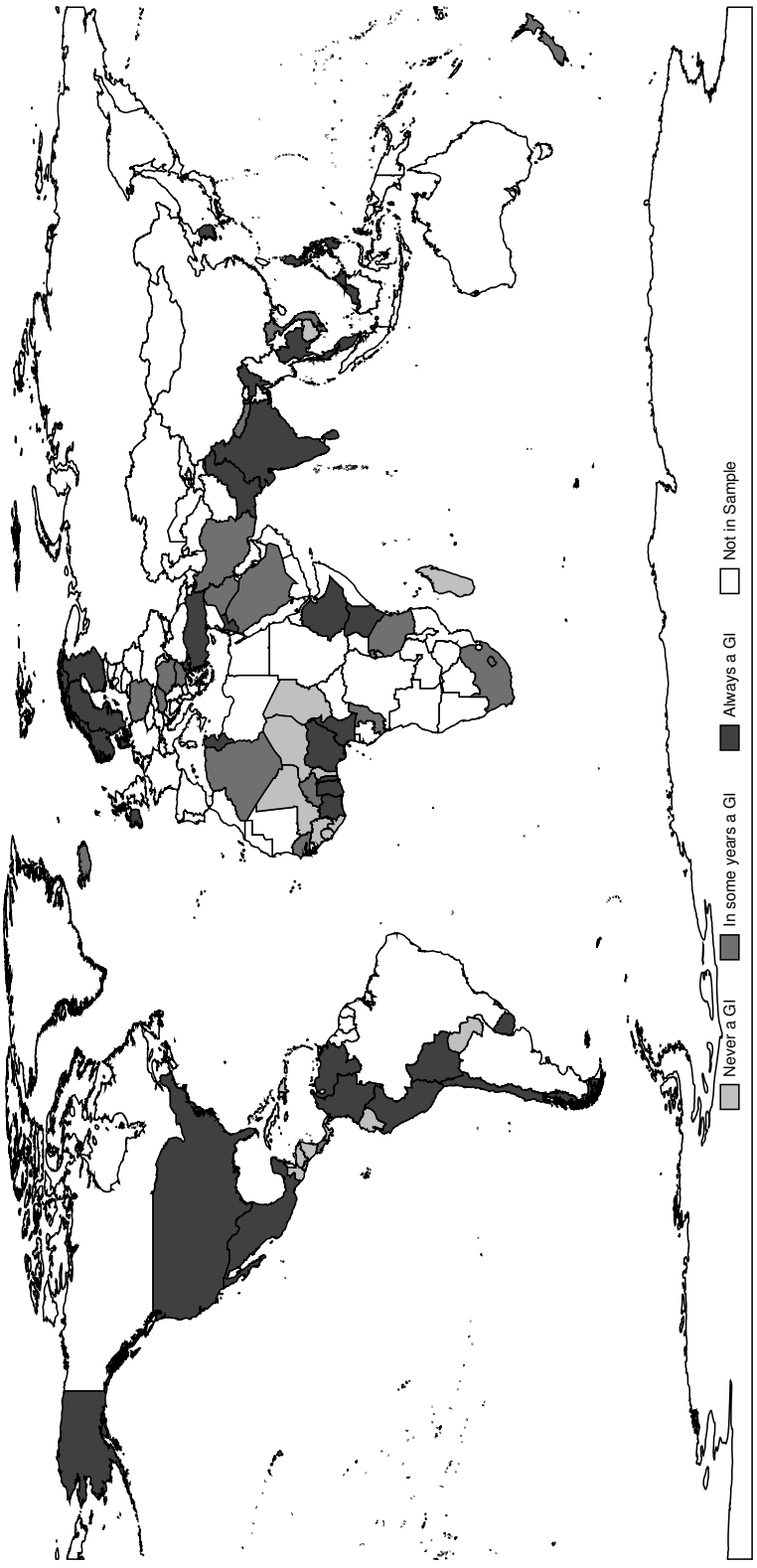


Figure A1: The presence of the GI.

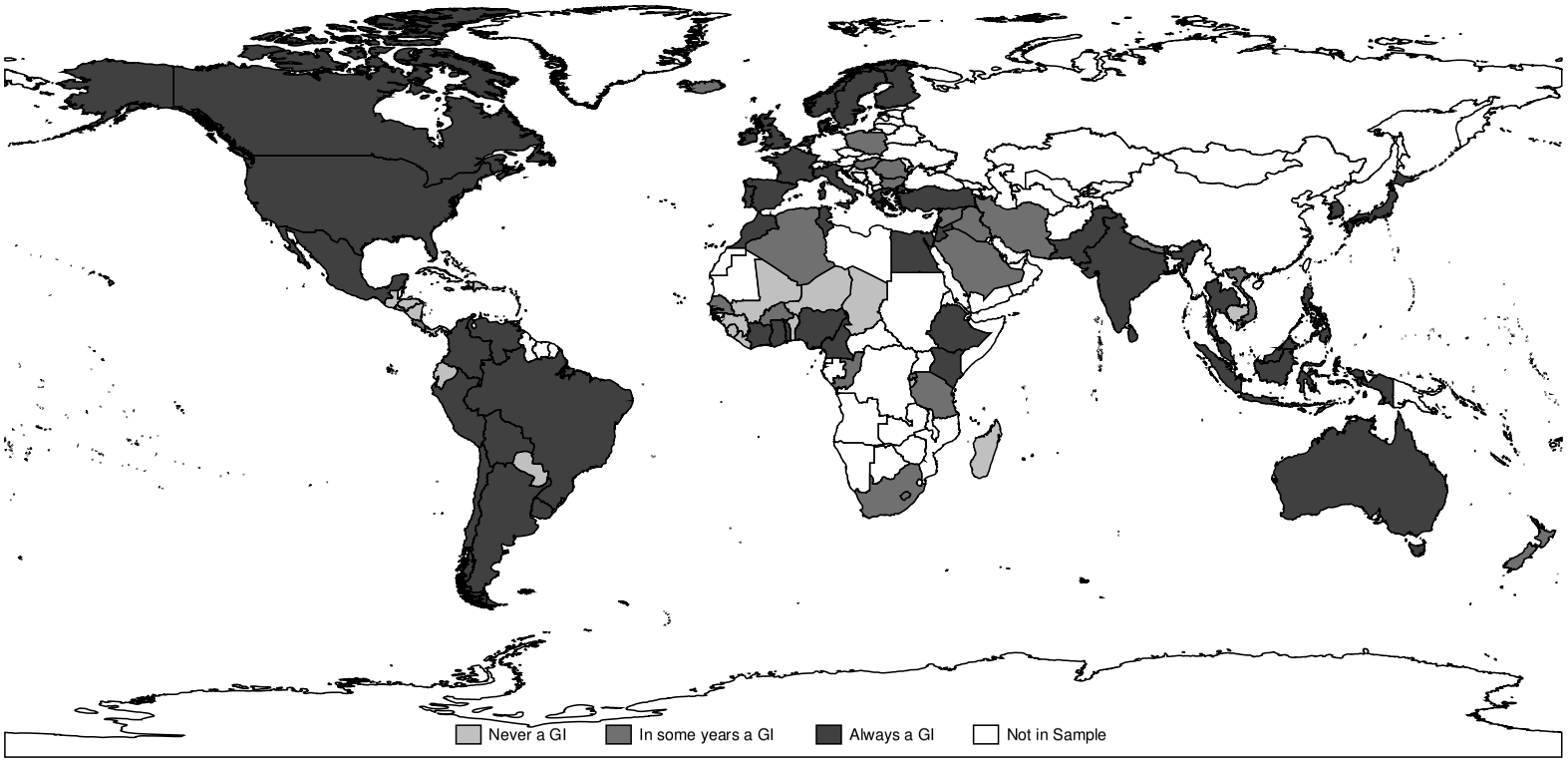


Figure A2: The presence of the GI - sample with jointly reported institutes.

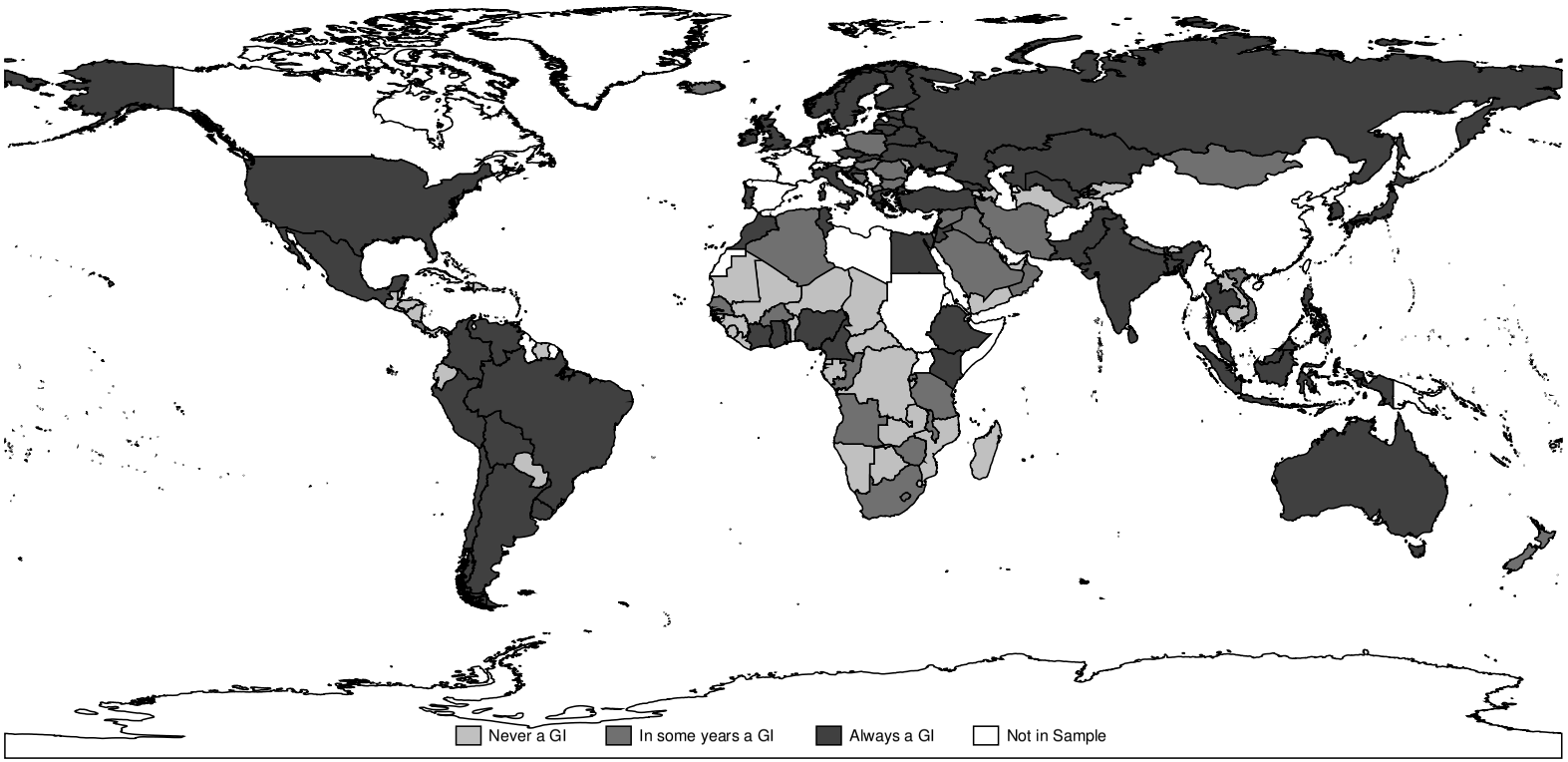


Figure A3: The presence of the GI - unbalanced sample.