Cultural similarities predict migration over and above shared location, shared language, and shared history

Extended abstract submitted to the European Population Conference 2022

Carolina Coimbra Vieira
coimbravieira@demogr.mpg.de
Max Planck Institute for Demographic Research
Rostock, Germany

Sophie Lohmann
lohmann@demogr.mpg.de
Max Planck Institute for Demographic Research
Rostock, Germany

Emilio Zagheni
zagheni@demogr.mpg.de
Max Planck Institute for Demographic Research
Rostock, Germany

ABSTRACT

One of the strongest empirical regularities in spatial demography is that flows of migrants are positively associated with population stocks at origin and destination, and inversely related to distance. This pattern, observed in the 19th century, was formalized into what are known as gravity models of migration. Traditionally, distance is measured geographically, however, other measurements including economic and cultural factors have also been found to be relevant to explaining migration flows. In particular, we believe that cultural distance may be one important form of distance to explain migration flows because it is dynamic. In this paper, we propose a scalable approach to obtain proxies for cultural similarity across countries by using data from the Facebook Advertising Platform. Our results show that our new measure of cultural similarity adds over and above standard predictors in predicting migration, opening new opportunities to understand determinants of migration.

1 INTRODUCTION

One of the strongest empirical regularities in spatial demography is that flows of migrants are positively associated with population stocks at origin and destination and inversely related to distance. This pattern was observed in the 19th century by Ravenstein [11] and later formalized by Zipf [19] into what are known as gravity models of migration. Traditionally, distance is measured geographically, however, other measurements including economic and cultural factors have also been found to be relevant to explaining migration flows [3, 8]. In particular, we believe that cultural distance may be one important form of distance to explain migration flows because it is dynamic.

Far from being static, migration is dynamically changing over time [4]. Standard forms of gravity models [5], in contrast, have typically included fairly static predictors such as the physical distance between two countries or whether countries share a colonial history. If migration patterns change, these standard predictors therefore cannot explain that change, whereas culture as a dynamic concept [6] would be better suited to explain that change. The cultural distance between two countries could therefore be a valuable predictor of migration flows. However, measures of cultural distance are difficult to estimate and thus have not yet been widely adopted in gravity models for assessing and predicting migration. We introduce a new measure of cultural distance that is quick, cost-effective, and scalable and show its value for explaining migration flows.

The few studies that have examined cultural distance in migration flows have typically relied on survey responses. For example, tourism flows into China illustrate that higher distance on cultural surveys is associated with less short-term mobility, over and above population stocks [17]. Similarly, immigration data from Denmark, Germany, and the Netherlands illustrate that higher distance on cultural surveys is also associated with less long-term mobility, again over and above population stocks [15]. However, these studies rely on survey data, such as the Hofstede's six cultural dimensions [17, 18] or the World Values Survey (WVS) [16]. Such data have multiple drawbacks: They are costly to collect, have limited sample sizes, and are subject to self-report response biases [13].

In this paper, we expand the methodology proposed by Vieira et al. (2020) [14] to propose a scalable approach to obtain proxies for cultural distance or cultural similarity across countries by using free data from the Facebook Advertising Platform (Facebook Ads). Moreover, we show that our new measure of cultural similarity explains migration flows over and above standard predictors and introduce a more nuanced view of symmetric and non-symmetric measures of cultural distance, opening new opportunities to understanding the determinants of migration.

2 DATA AND METHODOLOGY

Given the centrality of food as a cultural marker to study culture [2, 7], we selected 728 interests classified by Facebook as related to *Food and drink* from a dataset [12] containing most of the interests available on Facebook across 16 countries[1]. We obtained the number of monthly active Facebook users from each country who were interested in each of those 728 dishes. Based on these data, we identified the 50 most popular dishes in each country and constructed a vector indicator for each country based on the proportion of Facebook users interested in each dish. Finally, we constructed both symmetric and non-symmetric measures of cultural similarity. The symmetric measure (`CS_symm`) is given by the cosine similarity between the vectors of each pair of countries across all 728 interests. In other words, it describes how similar two countries' tastes in food are. The non-symmetric measures, in contrast, used only the top 50 dishes for any given country. For any pair of countries, there were thus two measures: For example, when predicting migration from Chile to Spain, we examined both the popularity of the top Chilean dishes in Spain (i.e., dishes from the country of origin; `CS_nonsymm_food_o`) and the popularity

[1] Argentina, Australia, Brazil, Chile, Great Britain, France, Indonesia, Japan, South Korea, Malaysia, Mexico, Russia, Singapore, Spain, Turkey, and the United States.

of Spanish dishes in Chile (i.e., dishes from the country of destination; `CS_nonsymm_food_d`). However, because symmetric and non-symmetric similarity measures were highly correlated, we examined them separately.

In addition to cultural similarity, our independent variables included the population (`pop_o`, `pop_d`) of each country in 2019[2]; the area (`area_o`, `area_d`) in km^2; the geographic distance between each pair of countries based on bilateral distances between the biggest cities of those two countries weighted by the share of the city in the overall country's population (`distwces`) between each pair of countries; the gross domestic product (`GDP_o`, `GDP_d`), GDP (constant 2010 US$) of each country in 2019[3]; an indicator of whether the two countries had ever had a colonial link (`shared_hist`) [9]; an indicator of whether the two countries shared a common official language (`col`) and linguistic proximity (`lp1`) between two countries [10]. Finally, the dependent variable corresponded to the migrant flow data (m_flow_da_pb_closed) by origin and destination estimated using the Demographic accounting, pseudo Bayesian approach for the period of 2015-2019[1].

We conducted a series of five gravity models: Model 1 refers to a simplified version of a gravity model [5] used as a baseline. Model 2 includes the GDP of both countries of origin and destination. Model 3 includes variables related to shared language and shared history. Model 4 and 5 add symmetric and non-symmetric cultural similarity, respectively, as additional predictors. Table 1 shows the resulting coefficients and statistics for each one of these models.

3 RESULTS

As shown in Table 1, the GDP of both countries of origin and destination are statistically significant and, when added to the gravity model, contributes to the increase of the adjusted r-squared from 0.08 (Model 1) to 0.40 (Model 2). Other variables related to shared language and shared history are also statistically significant and contribute to an adjusted r-squared of 0.53 (Model 3). The adjusted r-squared keeps increasing to 0.57 (Model 4) and 0.58 (Model 5) when our measures of cultural similarity are added to the gravity model. We observe that Model 4 and Model 5 have a similar adjusted r-squared, and both consider measures of cultural similarity. However, the interpretation of each one of them is slightly different. Model 4 has one variable representing cultural similarity describing how similar two countries' tastes in food are while Model 5 has two variables representing two measures of cultural similarity corresponding to the popularity of the dishes from the country of origin in the destination and vice-versa. Both models (Model 4 and Model 5) show that our measures of cultural similarity are good predictors of migration. However, the first four models (Model 1-4) consider symmetric variables to predict migration, which is clearly a non-symmetric measure. Since the migration flow between countries is non-symmetric (e.g., there are more Chilean in Spain than Spanish in Chile), we should expect that the cultural similarity in terms of food preferences will not be symmetric as well (e.g., there are more Chilean interested in Spanish food than Spanish interested in Chilean food). Because of that, we propose the use of a non-symmetric measure of cultural similarity in Model 5.

4 DISCUSSION

In this paper, we propose a new measure of cultural similarity and show how this measure can be used to explain migration flows between countries. Cultural similarity seems to hold an important role in predicting migration. Our results show that cultural similarity explains migration flows over and above standard predictors, such as shared language and shared history. While some variables such as shared language, history, and geographic distance are static or, at least, do not change easily, cultural attributes from our daily life are very sensitive to changes in the environment adding value to predictive models. Moreover, we introduce a more nuanced view of symmetric and non-symmetric measures of cultural distance, opening new opportunities to understanding the determinants of migration.

REFERENCES

[1] Guy J Abel and Joel E Cohen. Bilateral international migration flow estimates for 200 countries. *Scientific data*, 6(1):1–13, 2019.
[2] Bob Ashley, Joanne Hollows, Steve Jones, and Ben Taylor. *Food and cultural studies*. Routledge, 2004.
[3] Marcus H Böhme, André Gröger, and Tobias Stöhr. Searching for a better life: Predicting international migration with online search keywords. *Journal of Development Economics*, 2020.
[4] Francesco Castelli. Drivers of migration: why do people move? *Journal of travel medicine*, 2018.
[5] Joel E. Cohen, Marta Roig, Daniel C. Reuman, and Cai GoGwilt. International migration beyond gravity: A statistical model for use in population projections. 2008.
[6] Catherine R Cooper and Jill Denner. Theories linking culture and psychology: Universal and community-specific processes. *Annual review of psychology*, 1998.
[7] Isabelle De Solier and Jean Duruz. Food cultures: Introduction. *Cultural Studies Review*, 19(1):4–8, 2013.
[8] Victoria M Esses. Immigration, migration, and culture. In *Oxford Research Encyclopedia of Psychology*. 2018.
[9] Thierry Mayer and Soledad Zignago. Notes on cepii's distances measures: The geodist database. Working Papers 2011-25, CEPII, 2011.
[10] Jacques Melitz and Farid Toubal. Native language, spoken language, translation and trade. Working Papers 2012-17, CEPII, 2012.
[11] Ernest George Ravenstein. The laws of migration. *Journal of the royal statistical society*, 1889.
[12] Till Speicher, Muhammad Ali, Giridhari Venkatadri, Filipe Ribeiro, George Arvanitakis, Fabrício Benevenuto, Krishna Gummadi, Patrick Loiseau, and Alan Mislove. Potential for discrimination in online targeted advertising. In *Conference on Fairness, Accountability, and Transparency*, 2018.
[13] Edward A Suchman. An analysis of "bias" in survey research. *Public Opinion Quarterly*, pages 102–111, 1962.
[14] Carolina C. Vieira, Filipe N. Ribeiro, Pedro O.S.V. de Melo, Fabricio Benevenuto, and Emilio Zagheni. Using facebook data to measure cultural distance between countries: the case of brazilian cuisine. In *The Web Conference*, 2020.
[15] Roger White. Is Cultural Distance a Determinant of International Migration Flows? Evidence from Denmark, Germany, and the Netherlands. *Economics Bulletin*, 33(3), August 2013.
[16] Yang Yang, Hongbo Liu, and Xiang Li. The world is flatter? examining the relationship between cultural distance and international tourist flows. *Journal of Travel Research*, 58(2):224–240, 2019.
[17] Yang Yang and Kevin KF Wong. The influence of cultural distance on china inbound tourism flows: A panel data gravity model approach. *Asian Geographer*, 29(1):21–37, 2012.
[18] Yingsha Zhang, Xiang Li, and Tong Wu. The impacts of cultural values on bilateral international tourist flows: A panel data gravity model. *Current Issues in Tourism*, 22(8):967–981, 2019.
[19] George Kingsley Zipf. The p1 p2/d hypothesis: on the intercity movement of persons. *American sociological review*, 1946.

[2] https://www.un.org/en/development/desa/population/migration/data/estimates2/estimates19.asp
[3] https://databank.worldbank.org/home

	Model 1	Model 2	Model 3	Model 4	Model 5
(Intercept)	-5.74^{*}	-25.51^{***}	-25.64^{***}	-26.37^{***}	-25.31^{***}
	(2.42)	(2.65)	(2.38)	(2.27)	(2.24)
log10(pop_o)	0.33	-0.96^{***}	-0.55^{*}	-0.20	-0.09
	(0.26)	(0.28)	(0.25)	(0.25)	(0.26)
log10(area_o)	0.07	0.20^{*}	0.07	0.05	0.01
	(0.11)	(0.09)	(0.09)	(0.08)	(0.08)
log10(pop_d)	0.48	-1.19^{***}	-0.78^{**}	-0.43	-0.49
	(0.25)	(0.28)	(0.25)	(0.25)	(0.26)
log10(area_d)	0.16	0.32^{***}	0.19^{*}	0.17^{*}	0.14
	(0.11)	(0.09)	(0.09)	(0.08)	(0.08)
log10(distwces)	0.24	0.25^{*}	0.14	0.05	0.03
	(0.15)	(0.12)	(0.11)	(0.11)	(0.11)
log(GDP_o)		0.64^{***}	0.56^{***}	0.50^{***}	0.47^{***}
		(0.09)	(0.08)	(0.08)	(0.08)
log(GDP_d)		0.83^{***}	0.75^{***}	0.69^{***}	0.69^{***}
		(0.09)	(0.08)	(0.08)	(0.08)
col			1.12^{***}	0.49^{*}	0.36
			(0.22)	(0.24)	(0.25)
lp1			0.20^{***}	0.16^{***}	0.15^{***}
			(0.05)	(0.04)	(0.04)
shared_hist			0.63^{**}	0.49^{*}	0.48^{*}
			(0.23)	(0.22)	(0.22)
CS_symm				1.82^{***}	
				(0.36)	
CS_nonsymm_food_o					0.65^{***}
					(0.19)
CS_nonsymm_food_d					0.94^{***}
					(0.19)
R^2	0.10	0.41	0.55	0.59	0.60
Adj. R^2	0.08	0.40	0.53	0.57	0.58
Num. obs.	240	240	240	240	240

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 1: Resulting coefficients and statistics of the models to predict migration flows.