Growth Lab

Escaping from Hardship, Searching for Comfort: Climate Matching in Refugees' Destination Choices

Diego A. Martin, Jose Morales-Arilla, and Alvaro Morales

Growth Lab Working Paper Series
No. 237

October 2024

GROWTH LAB
HARVARD KENNEDY SCHOOL
79 JFK STREET
CAMBRIDGE, MA 02138

GROWTHLAB.HKS.HARVARD.EDU

Acknowledgments
We thank Ricardo Hausmann, Horacio Larreguy, Ed Glaeser, Dany Bahar, Sandra Rozo, Shreyas Gadgin Matha, Jesus Daboin, and participants at ITAM's Economic Research Seminar and at the Migration Dynamics Sessions at the Southern Economic Association's 2024 Conference for helpful comments and suggestions. Ruy Alleyne, Ana Ibarra, and Abigail Huie provided excellent research assistance. All errors are our own.

This paper may be referenced as follows: Martin, D.A., Morales-Arilla, J., Morales, A. (2024). "Escaping from Hardship, Searching for Comfort: Climate Matching in Refugees' Destination Choices." Growth Lab Working Paper, Harvard University Kennedy School of Government.

Escaping from hardship, searching for comfort: Climate matching in refugees' destination choices.*

Diego A. Martin[†] Jose Morales-Arilla[‡] Alvaro Morales [§]

October 29, 2024

[Word Count: 5,347]

Submitted AERI

Abstract

Do refugees settle in destinations that are ecologically similar to their origins? We assess the relevance of "climate matching" theories of migration for Venezuelan refugees in South America. Leveraging social media data, we build and validate the first local bilateral matrix of Venezuelan flows across the region. We measure bilateral ecological similarities in terms of temperature, precipitation, elevation, and distance to the coastline. Performing Poisson Pseudo-Maximum Likelihood gravity models of migration, we show that Venezuelan flows are more likely between ecologically similar areas. Model predictions explain independent measurements of Venezuelans' settlement choices at both bilateral and destination levels.

JEL Codes: F2, O15, R23

Keywords: Refugees, Mass migration, Climate matching, Gravity migration models, Social media.

*We thank Ricardo Hausmann, Horacio Larreguy, Ed Glaeser, Dany Bahar, Sandra Rozo, Shreyas Gadgin Matha, Jesus Daboin, and participants at ITAM's Economic Research Seminar and at the Migration Dynamics Sessions at the Southern Economic Association's 2024 Conference for helpful comments and suggestions. Ruy Alleyne, Ana Ibarra and Abigail Huie provided excellent research assistance. All errors are our own.

[†]Reskilling Lab, Harvard University, Cambridge, MA 02138. E-mail: dimartin@hbs.edu

[‡]Escuela de Gobierno y Transformación Pública, Tecnológico de Monterrey. E-mail: josemoralesarilla@tec.mx

[§]Harvard University, Cambridge, MA 02138. E-mail: amorales@hks.harvard.edu

1 Introduction

According to the UN, there were 117.3 million forcibly displaced individuals and 37.6 million refugees by the end of 2023, and this number has grown 44% since 2018 (UNHCR 2023). How does the growing number of forcibly displaced individuals choose where to settle after fleeing from conflict, natural disasters, and economic duress? The historical migration literature shows that destination choices are certainly multidimensional: Factors as diverse as economic prospects, coethnic presence, geographic proximity and political/cultural similarities seem highly influential. Similarly, ecological similarities between origins and destinations have been introduced in the literature as a determinant of historic mobility patterns: "Climate matching" theories of migration have long been proposed as explanations for European settlement patterns in North America (Fischer 1989) and South America (Lesser 2013, Devoto 2003), and these theories have been empirically validated in the context of the United States. (Obolensky et al. 2024).[1]

Does the ecological similarity between recent refugees' origins and their potential destinations influence their destination settlement choices? We evaluate this question in the context of the Venezuelan refugee shock across South America. Since 2017, about 7.7 million people (25% of the country's population) have left Venezuela, escaping the country's economic and political situation (IOM 2022), and 6.5 million of these refugees have stayed in South America.[2] We introduce a methodological framework to tackle this question in a way that is both locally precise and representative of the whole region, leveraging internationally comparable and precise measures of "bilateral" refugee flows and ecological similarities. Regarding the former, we build on the near universe of geolocalized tweets between 2011 and 2021 from Harvard's CGA Geotweet Dataverse (Lewis & Kakkar 2016).[3] We build the first matrix of subnational flows between Venezuelan and South

[1] Piguet (2013) noted that despite ignoring ecological factors as determinants of migration for a large part of the late XX Century, migration literature has come to consider the environment as part of the multidimensional set of variables that influence migration decisions. For instance, Graves (1980) proposes an amenities model for migration that considers the environment as a pull factor for movement decisions within the US. Along similar lines, Petersen (1958) sees nature as an ecological push for what he calls "primitive migration". Petersen refers to "conservative migrants" as those who look for similar ecological conditions to "seek only a place where they can resume their old way of life, and when this is possible, they are content."

[2] Venezuelans add to 5%, 4.5%, 2.5% and 2.4% of the population in Colombia, Peru, Chile and Ecuador, respectively (UNHCR 2023). When compared with the international response to recent refugee crises in Syria and Ukraine, the support given to developing South American countries bearing the brunt of the Venezuelan exodus has been lackluster (Bahar & Dooley 2019).

[3] Social media offers a new way to measure migration and refugee flows. Using the geolocated activity of individual accounts on social media platforms, it is possible to "observe" the movements of certain individuals across space (Drouhot et al. 2022). People who permanently start producing or consuming social media content in areas other than where they initially engaged with such content earlier can be considered migrants (Spyratos et al. 2018). The problem of tracking migration with social media comes from the lack of representativeness of the resulting measures. When a migrant group does not use social media, it is impossible to calculate their flows between different locations (Olteanu et al. 2019). For instance, Hübl et al. (2017) show that using geo-tagged tweets to identify refugee migration patterns from the Middle East and Northern Africa to Europe has limitations since only a small portion of refugees use Twitter. Still, prior exercises using geolocated Social Media data have helped predict independent measures of migrants' outflows and inflows (Zagheni et al. 2014). For 17 EU countries, Spyratos et al. (2018) employed Facebook information to estimate the number of "expats" in each. Böhme et al. (2020) use geo-referenced online search data in the origin country to predict migration to OECD countries. Most importantly, geolocated Twitter data has been used effectively to detect precise destinations of Venezuelan migrants across South America (Mazzoli et al. 2020, Hausmann et al. 2018).

American localities by counting the number of accounts based on a specific Venezuelan origin between 2011 and 2016 that were based in a South American destination outside Venezuela between 2017 and 2021. We validate this measurement by showing that the user flows identified in the matrix correlate with independent measures of the refugee shock at both bilateral level and destination levels.[4] While there are no independent local measures of emigration at the origin level that we are aware of, we validate our Twitter-based flow matrix at the origin-level with electoral data from 2024: A Twitter-based emigration index associates with a drop in electoral turnout and with an increase in regime support.[5]

To test the "climate matching" hypothesis in the context of the Venezuelan refugee shock, we need to complement our refugee flow matrix with bilateral measures of the ecological similarities between each pair of Venezuelan origins and South American destinations. For this purpose, we leverage the "GLocal" dataset (Morales-Arilla & Gadgin Matha 2024) to collect local measures of the average temperature, average precipitation, distance to the coastline and elevation for all second-level administrative units in South America. We measure bilateral similarities for each feature as the inverse of the standardized absolute difference between each Venezuelan origin and each South American destination. Moreover, we measure bilateral population density similarities and driving distances between all origin-destination pairs leveraging the methodology proposed in Luxen & Vetter (2011).[6]

Building on a Poisson Pseudo Maximum Likelihood (PPML) specification, we perform gravity models that predict bilateral refugee flows as a function of ecological similarities between local origins and destinations.[7] Our results suggest that Venezuelan refugees moved to areas with similar temperatures, precipitation levels, and coastline distances as observed in their Venezuelan origins. These results are robust in controlling for the driving proximity between localities and their population-density similarity. Moreover, to address potential concerns about non-classical measurement error in our outcome variable, we replicate our analyses by restricting the sample to non-rural pairs, non-capital pairs, and flows outside Venezuelan or Colombian/Brazilian border areas. We show that our estimates are robust to these sample choices in the Online

[4] To the extent of our knowledge, independent origin-destination level data on the Venezuelan refugee shock is only available in the Colombian context, and comes from the *"Pulso de la Migración"* Survey. Similarly, detailed administrative aggregates on the destination of Venezuelan migrants are only publicly available in the Colombian context and are provided by the country's migration authority (*Migración Colombia*).

[5] These results are expected for an adequate index of migration at the origin level, as emigrants remained registered to vote but became unable to cast their vote from abroad, and because opponents are more likely to have chosen to migrate in the years prior to the 2024 election (Hirschman 1970).

[6] We thank Shreyas Gadgin Matha for his help in producing these estimates.

[7] Given the bilateral nature of migration, the "gravity" empirical methodology developed in the context of international trade has been extended to the context of migration. Gravity models are a helpful tool for understanding the determinants of bilateral migration between countries (Ramos 2016, Beine et al. 2016). Gravity models have explained that migration grows with a rise in the income of destination communities (Docquier 2018, Rikani & Schewe 2021) and a decline in the environmental conditions in the origin location (Maurel & Tuccio 2016, Cai et al. 2016, Helbling & Meierrieks 2021). Recently, Beyer et al. (2022) have shown that the gravity models often fail to capture temporal migration dynamics in panel data contexts.

Appendix.[8] Finally, we show that predictions from our "climate matching" gravity model of the Venezuelan refugee shock help explain independent measures of Venezuelans' settlement choices.

This paper contributes to the migration and refugees literature in two distinct ways: First, it provides and validates the first local bilateral matrix of individual flows in the context of the Venezuelan refugee shock across South America - one of the largest human exoduses in recent history. This matrix, and its aggregations at the origin or destination levels, can help inform the study of the causes and consequences of the Venezuelan refugee shock, both domestically and regionally.[9] Most importantly, our results indicate that "climate matching" theories of migration that have been proposed and tested in historical settings are relevant to explain individual settlement choices during recent, massive refugee shocks.[10] Venezuelan refugees seem to settle in destinations that are ecologically similar to their domestic origins, and predictions from our ecological similarity based model help identify the areas where Venezuelans eventually settled in. Unfortunately, increasingly prevalent ecological, political and economic shocks are pushing the number of refugees around the World upwards. Understanding that ecological similarities inform refugee settlement choices can help researchers and policy-makers better target resources and attention towards areas that are ecologically similar to those that experience such shocks in the future.

The paper continues as follows: First, we describe the calculation of our bilateral measures of Venezuelan refugee flows, reporting the results for validation exercises at the bilateral, destination and origin levels. Then, we describe how we measure the ecological similarities between local Venezuelan origins and destinations in other South American countries. Third, we describe and report our gravity models of Venezuelan

[8] The main concern is that, due to the fact that measuring bilateral flows with Social Networking data requires connections to telecommunication networks in both origins and destinations, the method may be missing flows from/to connected to/from disconnected areas. If connected areas are ecologically similar, we might only observe a correlation between user flows and ecological similarity due to outcome measurement problems. To address this, Table A.2 shows that our main results are robust to removing rural areas -where telecommunication access constraints are concentrated- and to removing capital cities -where telecommunication access should be greatest-. Moreover, flows to and from border towns may be especially prevalent, as borders are largely porous. For this reason, our method may be over-counting flows between ecologically similar border towns that do not relate to permanent refugee flows. To assess whether our findings are driven by this possibility, Table A.3 excludes either Venezuelan border towns or Colombian/Brazilian border towns from the analysis, and shows that our main findings remain.

[9] To the extent of our knowledge, there are no quantitative empirical studies focusing on the domestic causes of the Venezuelan refugee shock, and only one study that studies the domestic consequences: Sviatschi et al. (2024) find that migration led to increases in local inequality and in regime support. While better established, the literature on the political and economic consequences of the Venezuelan refugee shock across South America has concentrated on the Colombian setting (Ibáñez et al. 2024, Rozo & Vargas 2021, Bahar et al. 2021, Ibáñez et al. 2021), with much fewer contributions focusing on other South American countries (Ajzenman et al. 2023, Groeger et al. 2024, Ryu & Paudel 2022, Olivieri et al. 2022, Marques 2024). While Colombia has faced the brunt of this refugee shock (World Bank 2018), the specificity in the country's policy openness to Venezuelans, its unique experience with displacement, and the cultural and historical connection between the two countries makes it hard to generalize findings to the rest of the region.

[10] While limited by specific circumstances, studies show that refugees often retain substantive agency in their settlement choices. Beyond economic opportunity and geographic proximity, refugees often settle in localities to which they have historical, cultural and ethnic ties (Moore & Shellman 2007, Zorlu & Mulder 2008, Connor 1989, Mossaad et al. 2020, Martén et al. 2019, Zavodny 1999). Different studies have found these factors to be relevant for the specific context of the Venezuelan refugee shock (Ayala et al. 2020, Pirovino & Papyrakis 2023, Díaz-SáNchez et al. 2021, Galindo et al. 2022, Saa et al. 2020). While there is a dense literature on the existence of "climate refugees" (Piguet 2013, Beine & Parsons 2017, McGregor 1994, Black 1994), it largely focuses on climate change and ecological shocks as push factors forcing refugees from their homes, and not on settlement considerations of environmental characteristics of potential destinations in relation to those of their origins. To the extent of our knowledge, this is the first paper to consider ecological similarity as a key determinant of refugee's settlement choices.

refugee flows as a function of bilateral ecological similarities, and show how model predictions help estimate independent measures of Venezuelans' settlement choices. Finally, we present the paper's conclusions.

2 Bilateral local migration of Venezuelan refugees across South America

We measure the local bilateral flows between Venezuela and other South American Countries with social networking data from Twitter.[11] Harvard's Center for Geographic Analysis (CGA) keeps a collection of the near universe of geolocated tweets since 2010.[12] We take this data to identify the Venezuelan origins of different Twitter users before the start of the refugee crisis in 2017, and then determine their destination after the start of the Venezuelan refugee shock. We assign origins and destinations to different users with according to the following regularity and concentration protocol:

1. We take data for the baseline period 2013-2016 and identify users that posted tweets from Venezuela. We identify the administrative area for tweets connected to these users.
2. We filter users to keep those that have at least 10 baseline tweets within Venezuela, with a majority of them coming from a single Venezuelan administrative unit. We assign this unit as the user's origin.
3. We now take tweets from these users for the endline period 2017-2020, and keep only those that have endline tweets in other South American country.
4. We filter this list to keep only users with at least 5 endline tweets in South America, and a majority of them coming from the same administrative unit. We assign this unit as the user's destination.

After implementing this protocol, we identify 64,036 users with a specific Venezuelan origin and South American destination. We count the number of users at each origin and destination combination to get our Twitter-based matrix of bilateral flows. Panels A and B of figure 1 show the localities of South American with one or more migrants from the Caracas and the Caroní municipalities in Venezuela. This matrix is both sparse and skewed: First, only about 0.6% of OD combinations have non-zero flows. Second, one combination of the 2.7MM possible origins and destinations (Caracas to Buenos Aires) captures 1,532 flows, 2.4% of all flows in the data. Moreover, the fact that the top destination administrative unit in our data is

[11] In all our analyses, we use the administrative unit definitions used in the GLocal dataset. These come from the GADM 3.6 definition (Hijmans et al. 2018). Importantly, we focus on the second administrative level in all South American countries (e.g. municipalities) as our "locality" definition. We consider the first administrative level for all countries as the definition of a "state", which we later use as regression controls and for clustering of standard errors.

[12] There are two types of geolocations that users can provide: A tweet-specific geolocation of the originating device, and a user-specific geolocation of a designated place of origin. Importantly, the latter may change for tweets of the same user if they change the designated place. According to Harvard's CGA, between 1% and 2% of all tweets contain at least one of these two geolocations. Whenever both measures are available, we consider the tweet-specific geolocation.

Buenos Aires (a distant location chosen by relatively well-off migrants) points to the biases inherent in using social networking data to measure refugee flows. In order to validate our Twitter matrix of bilateral flows as an informative measure of the Venezuelan refugee shock, we evaluate whether it correlates with independent measures of the refugee shock at both the bilateral and destination levels. We do this in the context of Colombia, which is the only country for which there is both bilateral and location-specific data about the arrival of Venezuelan refugees.

At the bilateral level, we run the following regression:

$$F_{o,d} = \beta_0 + \beta_1 T_{o,d} + \beta_2 P_{o,d} + \phi_o + \phi_d + \epsilon_{o,d} \tag{1}$$

Where $F_{o,d}$ is the inverse hyperbolic sine (IHS)[13] of the number of Venezuelans from origin location o in destination location d detected in Colombia's *Pulso de la Migración* Survey.[14] $T_{o,d}$ is the IHS of the Twitter user flows from origin o in destination d. $P_{o,d}$ is the IHS of the total population in the o and d dyad in 2005.[15] ϕ_o and ϕ_d stand for origin and destination fixed effects, and $\epsilon_{o,d}$ is an error term. A positive and significant value of β_1 would suggest that both measures of bilateral flows are correlated after controlling for scale factors.

At the destination level, we perform the following specification:

$$F_d = \beta_0 + \beta_1 T_d + \beta_2 P_d + \phi_s + \epsilon_d \tag{2}$$

Where F_d is the IHS of the number of Venezuelans in destination locality d as measured in Colombian administrative data.[16] T_d is the IHS of the number of Twitter users from Venezuelan origins identified in destination d. P_d is the IHS of the population of Colombian destination d. ϕ_s is a State-level fixed effect, and ϵ_d is the error term. A positive and significant value of β_1 would suggest that both measures of destination-level flows are correlated after controlling for scale factors.

Table 1 provides estimates for Equation 1 and Equation 2. Standard errors are estimated allowing for

[13] We use the inverse hyperbolic sine transformation in these validation analyses to deal with both the sparsity and skewedness of the data with a cogent approach across both bilateral and destination levels. For more on the use of the use of the inverse hyperbolic sine transformation, see Bellemare & Wichman (2020).

[14] To the extent of our knowledge, Colombia's *Pulso de la Migración* survey is the only instrument in the region that captures local origins and destinations of Venezuelan migrants. We take data from the 2022 survey and aggregate expansion factors at the origin-destination levels to get an independent measure of bilateral flows.

[15] All local population measures used in this study are taken for 2005 from the GLocal dataset. Its original source was NASA's Gridded Population of the World data. We also take locality areas to build 2005 measures of population density.

[16] Colombia's migration authority identifies and publishes local aggregates of the destination locations of Venezuelan migrants in the country. We take this data point for 2022 as an independent measure of the local destination of Venezuelans in Colombia.

error correlation within State combinations or within State levels. Results suggest that our Twitter-based flow matrix correlates with independent measures of the Venezuelan refugee shock: Columns 1 and 2 focus on the bilateral level. Column 1 shows that a 1% increase in the number of Twitter flows associates with a 0.6% increase in the number of flows in the *Pulso de la Migración* Survey. Column 4 shows that, while attenuated, this relationship is robust to controlling for the aggregate population size of each dyad. Columns 3 and 4 focus on destination-level aggregates. Column 3 shows that a 1% increase in the number of Twitter flows into a Colombian municipality associates with a 0.8% increase in the number of Venezuelans in a municipality according to Colombian administrative data. Column 4 again shows that, while attenuated, the direction and significance of this estimate is robust to controlling for municipalities' population. These findings confirm that our Twitter-based matrix of Venezuelan flows provides informative data about the Venezuelan refugee shock at the bilateral and destination levels.

Unfortunately, there is no independent measure of local Venezuelan emigration that we can contrast our Twitter-based flow matrix to. For this reason, we validate our matrix at the origin level by assessing its association with changes to the Electoral turnout and regime support rates in Venezuelan presidential elections. The Venezuelan opposition started to boycott electoral events in 2017, coinciding with the peak of the refugee shock.[17] However, the opposition again participated in the presidential elections of July 28th, 2024. The 5 million eligible voters that left the country were unable to update their registration in order to vote from abroad (The World 2024). While the electoral authority published an aggregate result for the election marking President Nicolas Maduro as the winner, it did not release detailed information at the voting booth level. The opposition collected official printed tallies for 83% of the country's voting booths, publishing scanned PDFs of the tallies and a dataset with booth-specific results and participation levels. These tallies indicate that Edmundo González -the opposition candidate- was the clear winner. Independent audits of the opposition-published results point to their veracity (Kronick 2024), which implies that the official aggregate results are fraudulent.[18]

We take official data on electoral turnout and regime support rates by municipality for the 2012 and 2013 Presidential elections and the 2015 Parliamentary elections, and complement it with municipality-level aggregates from the 2024 opposition-published results. Importantly, there are 11 out of 335 municipalities for which the opposition was unable to retrieve and publish any booth-level tallies. We exclude these municipalities from the analysis. We build an origin-level migration index by standardizing the ratio between the number of Twitter user flows from each Venezuelan municipality and the respective number of registered voters (electors) in 2024. Panels A and B of Figure 2 show the total number of emigrants and the migration

[17] Details about Venezuela's political context during this period can be found in Morales-Arilla (2024).

[18] For more details on the specifics of the 2024 elections, see Caracas Chronicles (2024).

index of municipalities for which there is available 2024 electoral data. We run the following TWFE difference-in-differences regression model:

$$E_{m,y} = \alpha + \sum_{y \neq 2015} \beta_y \times M_m \times 1[Y = y] + \phi_y + \phi_m + \epsilon_{m,y} \tag{3}$$

Where $E_{m,y}$ is the electoral outcome of interest (electoral turnout or regime support rate) in municipality m in election y, M_m is the migration index for municipality m, ϕ_y and ϕ_m are election and municipality fixed effects, and $\epsilon_{m,y}$ is an error term. Standard errors are calculated allowing for within-municipality correlation of errors, and observations are weighted by the number of 2024 electors. We use the parliamentary election of 2015 as a reference election, so that β_y captures the difference in the association between the migration index and the electoral outcome of interest between election y and the 2015 election. If our Twitter-based migration index is an adequate measure of migration, we would expect for it to capture a decrease in electoral turnout in 2024 with respect to all prior elections. This is because refugees remained registered to vote in their municipalities of origin, but became unable to cast their vote. Moreover, observing that regime support was relatively resilient in 2024 in comparison to prior elections in areas with high migration indices would also serve to validate the index if migrants were more likely to oppose the regime, which would be expected from the perspective of Hirschman (1970).[19] Hence, we expect β_{2024} to be negative and significant when focusing on electoral turnout rates, and for it to be positive and significant when focusing on regime support rates. In both specifications, we test for the plausibility of the parallel trend assumption by checking whether the estimates for β_{2012} and β_{2013} are smaller and statistically indistinguishable from 0.

Panels C and D of Figure 2 show that the expectations described above do hold in the data. A 1 standard deviation increase in our Twitter-based migration index associates with about 2pp decrease in the electoral turnout rate, and a similar increase in the regime support rate, with respect to 2024. All baseline estimates are much smaller and statistically insignificant. Now again, these results provide evidence in favor of the validity of our Twitter-based flow matrix when aggregated at the origin level.

3 Bilateral ecological similarities

We take comparable information on ecological characteristics of different localities from the "GLocal" dataset (Morales-Arilla & Gadgin Matha 2024). We focus on four key ecological characteristics: Average Precipita-

[19] This would be consistent with the findings discussed in Sviatschi et al. (2024), who approximate the pull-factor for return Colombian migrants with the baseline Colombian presence in different Venezuelan localities.

tion, Average Temperature, Elevation and Distance to the coastline.[20] We calculate the similarity between origins and destinations for each of these measures as the inverse of their standardized absolute difference.[21] We also standardize and invert the driving distances between all local origin and destination dyads. Finally, we calculate a similar measure of the bilateral similarity in population density in 2005. Panels C and D of Figure 1 show the Temperature Similarity measures between the Caracas and the Caroní municipalities and all other localities across South America. Summaries of all variables used in this study at the bilateral, destination and origin levels are in Table A.1.

4 Testing for climate matching in Venezuelan refugees' settlement choices

We assess the relevance of climate matching theories of migration in the context of the Venezuelan refugee shock by running a Poisson Pseudo-Maximum Likelihood gravity model of bilateral flows in our Twitter-based matrix as a function of the ecological similarities mentioned above. More specifically, we perform the following specification:

$$T_{o,d} = \exp(\beta_0 + \sum_{v=1}^{4} \beta_v S^v_{o,d} + \sum_{c=1}^{C} \beta_c X^c_{o,d} + \phi_o + \phi_d + \epsilon_{o,d}) \tag{4}$$

Where $T_{o,d}$ is the total number of Twitter user flows from origin o in destination d, $S^v_{o,d}$ is one of the four v ecological similarity measures of interest, $X^c_{o,d}$ are a set of C bilateral controls, ϕ_o and ϕ_d are origin and destination fixed effects, and $\epsilon_{o,d}$ is an error term. We expect the value of a specific β_v estimate to be positive and significant if the ecological similarity between origins and destinations on characteristic v is relevant for Venezuelan refugees' settlement choices. We calculate standard errors allowing for within State combination correlation in errors.

Table 2 provides estimates for this regression specification. Columns 1 to 4 test for the relevance of the different characteristics separately. Bilateral similarities in terms of Rainfall, Distance to the Coastline and Temperature seem to influence the number of Twitter flows between origins and destinations, while Elevation similarity shows a statistically insignificant estimate. Column 5 confirms these patterns when all similarities are considered jointly as described in Equation 4. The positive and statistical significance of Rainfall, Temperature and Distance to Coastline Similarities are robust to controlling for the driving

[20] Original precipitation and temperature data aggregated in the GLocal dataset are from the Global Precipitation Climatology Project (GPCP). We average monthly aggregates for the period 2010-2020. Original elevation data comes from the United States' Geological Service. Distances to the coastline are calculated as the geodesic distance between the population-weighted centroid of each locality to the coastline shapefile from Natural Earth (2024).

[21] We first calculate absolute differences in the measure of each feature between each origin and destination combination. For the purpose of comparability of regression estimates across measures, we then standardize all these differences. Finally, for ease of interpretation as estimates of ecological similarities, we multiply all standardized differences by -1.

proximity between origins and destinations (Column 6), Population Density similarity (Column 7) or both (Column 8). Interestingly, estimates of the effect of Elevation similarity turn negative and significant in model regressions that control for the bilateral driving proximities (Columns 6 and 8). Importantly, since all predictor variables are standardized, we can compare the relative importance of a 1 standard deviation increase in the proximity of different variables between them. For instance, the importance of the similarity in the distance to the coastline seems to be about half that of the bilateral driving proximity between origin and destinations, and is more than an order of magnitude superior to the importance of population density similarities. While somewhat smaller, similar comparisons can be drawn for rainfall and temperature similarities. Finally, the negative estimate of elevation similarity seems much smaller in absolute value than the positive values estimated for the other ecological similarity measures. Table A.2 shows that these results are largely robust for samples that exclude either rural pairs or capital pairs. This suggests that results are not driven by outcome measurement problems in areas with either very high or very low access to telecommunication technologies. Table A.3 shows results are robust to excluding flows from or to border towns, which are likely to be very similar just because of their proximity to each other.

Beyond finding positive and significant estimates of β_v for three of our four ecological similarity measures, we want to assess whether gravity models that explain Twitter flows strictly as a function of ecological similarities can make informative predictions about the settlement choices of Venezuelan refugees. For this purpose, we take the model of Column 5 in Table 2, which only has ecological similarity variables in its right hand side. We calculate bilateral fitted values based strictly on the coefficients placed on the different ecological similarity measures. *That is, we exclude the origin and destination fixed effects from the prediction.* Finally, we exponentiate these fitted values to get PPML ecological predictions of the number of flows between each locality dyad.

To test whether these predictions are informative of Venezuelans' settlement choices, we perform similar regression specifications as those described in Equations 1 and 2, but using our model predictions and their aggregation at the destination level. Table 3 shows a positive association between model predictions and independent measures of the Venezuelan refugee shock. Columns 1 and 2 focus on the bilateral level, and show that a 1% higher number of predicted bilateral Twitter flows associate with a 0.2% higher number of Venezuelans in the *Pulso de la Migración* Survey. Columns 3 and 4 focus on the destination level, and show that a 1% increase in the number of predicted Twitter flows to a given locality associate with a proportional increase in the number of Venezuelans in that locality according to Colombian administrative data. Taken together, these results suggest that gravity models based on ecological similarities between origins and destinations can help predict refugees' settlement choices.

5 Conclusion

This paper makes two key contributions to our understanding of refugee settlement patterns in general, and for the Venezuelan refugee crisis in particular. First, we introduce and validate a novel methodology for measuring subnational bilateral refugee flows using social media data. By analyzing geolocated Twitter activity between 2011 and 2021, we construct the first comprehensive origin-destination matrix of Venezuelan refugee flows across South America. Our validation exercises demonstrate that these social media-derived measures correlate significantly with independent data on refugee settlements and produce theoretically consistent effects on domestic political outcomes in Venezuela. Second, we provide empirical evidence that "climate matching" theories of migration, previously studied primarily in historical contexts, help explain contemporary refugee settlement patterns. Our gravity model specifications reveal that Venezuelan refugees systematically chose destinations with similar ecological characteristics to their origins – particularly in terms of temperature, precipitation, and coastal proximity. These findings hold robust across various sample specifications and when controlling for other determinants of migration such as geographic distance and population density similarities. Importantly, our model's predictions successfully explain independent measures of Venezuelan settlement patterns in Colombia, suggesting that ecological similarity serves as a meaningful predictor of refugee destination choices.

These findings have important implications for both academic research and policy design. For researchers, our methodology demonstrates the potential of social media data to generate detailed bilateral migration measurements in contexts where conventional data collection is challenging. The validated origin-destination matrix we construct can serve as a valuable input for future studies examining both the causes and consequences of the Venezuelan exodus across South America. For policymakers, our results suggest that ecological similarities between origin and destination locations can help predict where refugees are likely to settle during mass displacement events. This insight could improve the targeting of humanitarian resources and integration support services in future refugee crises, particularly as ecological, political, and economic shocks continue to drive forced displacement worldwide. Results also point to promising avenues for future research. Investigating the importance of ecological similarity in refugee settlement choices across populations and contexts could help assess the generalizability of our results. Moreover, understanding how ecological similarities interact with other relevant factors could provide a more complete understanding of refugee settlement choices. Finally, studying if better climate matching influences refugees' integration outcomes could help inform refugee resettlement policies.

6 References

Ajzenman, N., Dominguez, P. & Undurraga, R. (2023), 'Immigration, crime, and crime (mis) perceptions', *American Economic Journal: Applied Economics* **15**(4), 142–176.

Ayala, Y., Bayona, J. A., Karaeminogullari, A., Perdomo-Ortíz, J. & Ramos-Mejía, M. (2020), 'We are very similar but not really: the moderating role of cultural identification for refugee resettlement of venezuelans in colombia', *Frontiers in Psychology* **11**, 569394.

Bahar, D. & Dooley, M. (2019), 'Venezuela refugee crisis to become the largest and most underfunded in modern history', *Brookings Institution* **9**.

Bahar, D., Ibáñez, A. M. & Rozo, S. V. (2021), 'Give me your tired and your poor: Impact of a large-scale amnesty program for undocumented refugees', *Journal of Development Economics* **151**, 102652.

Beine, M., Bertoli, S. & Fernández-Huertas Moraga, J. (2016), 'A practitioners' guide to gravity models of international migration', *The World Economy* **39**(4), 496–512.

Beine, M. & Parsons, C. R. (2017), 'Climatic factors as determinants of international migration: Redux', *CESifo Economic Studies* **63**(4), 386–402.

Bellemare, M. F. & Wichman, C. J. (2020), 'Elasticities and the inverse hyperbolic sine transformation', *Oxford Bulletin of Economics and Statistics* **82**(1), 50–61.

Beyer, R. M., Schewe, J. & Lotze-Campen, H. (2022), 'Gravity models do not explain, and cannot predict, international migration dynamics', *Humanities and Social Sciences Communications* **9**(1), 1–10.

Black, R. (1994), 'Forced migration and environmental change: the impact of refugees on host environments', *Journal of Environmental Management* **42**(3), 261–277.

Böhme, M. H., Gröger, A. & Stöhr, T. (2020), 'Searching for a better life: Predicting international migration with online search keywords', *Journal of Development Economics* **142**, 102347.

Cai, R., Feng, S., Oppenheimer, M. & Pytlikova, M. (2016), 'Climate variability and international migration: The importance of the agricultural linkage', *Journal of Environmental Economics and Management* **79**, 135–151.

Caracas Chronicles (2024), 'Anatomy of a fraud'. Accessed: 2024-10-12.
URL: https://www.caracaschronicles.com/2024/09/04/anatomy-of-a-fraud/

Connor, K. M. (1989), 'Factors in the residential choices of self-settled afghan refugees in peshawar, pakistan', *International Migration Review* **23**(4), 904–932.

Devoto, F. (2003), 'Historia de la inmigración en la argentina', *(No Title)* .

Díaz-SáNchez, J. P., Bonilla-Bolaños, A. & Obaco, M. (2021), 'Where to go? drivers of venezuelan asylum seekers', *Journal of Refugee Studies* **34**(2), 1713–1729.

Docquier, F. (2018), 'Long-term trends in international migration: Lessons from macroeconomic model', *Economics and Business Review* **4**(1), 3–15.

Drouhot, L. G., Deutschmann, E., Zuccotti, C. V. & Zagheni, E. (2022), 'Computational approaches to migration and integration research: promises and challenges'.

Fischer, D. H. (1989), *Albion's seed: Four British folkways in America*, Vol. 1, America: A Cultural History.

Galindo, G., Navarro, J., Reales, J., Castro, J., Romero, D., Rodriguez A, S. & Rivera-Royero, D. (2022), 'Immigrants resettlement in developing countries: A data-driven decision tool applied to the case of venezuelan immigrants in colombia', *Plos one* **17**(1), e0262781.

Graves, P. E. (1980), 'Migration and climate', *Journal of regional Science* **20**(2), 227–237.

Groeger, A., León-Ciliotta, G. & Stillman, S. (2024), 'Immigration, labor markets and discrimination: Evidence from the venezuelan exodus in perú', *World Development* **174**, 106437.

Hausmann, R., Hinz, J. & Yildirim, M. A. (2018), Measuring venezuelan emigration with twitter, Technical report, Kiel working paper.

Helbling, M. & Meierrieks, D. (2021), 'How climate change leads to emigration: Conditional and long-run effects', *Review of Development Economics* **25**(4), 2323–2349.

Hijmans, R., Garcia, N. & Wieczorek, J. (2018), 'Gadm: database of global administrative areas, version 3.6', *GADM Maps and Data* .

Hirschman, A. O. (1970), 'Exit, voice, and loyalty: Responses to decline in firms, organizations, and states'.

Hübl, F., Cvetojevic, S., Hochmair, H. & Paulus, G. (2017), 'Analyzing refugee migration patterns using geo-tagged tweets', *ISPRS International Journal of Geo-Information* **6**(10), 302.

Ibáñez, A. M., Moya, A., Ortega, M. A., Rozo, S. V. & Urbina, M. J. (2024), 'Life out of the shadows: the impacts of regularization programs on the lives of forced migrants', *Journal of the European Economic Association* p. jvae044.

Ibáñez, A. M., Rozo, S. V. & Urbina, M. J. (2021), 'Forced migration and the spread of infectious diseases', *Journal of Health Economics* **79**, 102491.

IOM (2022), Venezuelan situation, Technical report, International Organization for Migration—IOM.

Kronick, D. (2024), 'On the validity of vote counts published by the venezuelan opposition'. Accessed: 2024-10-12.
URL: https://dorothykronick.com/28J.pdf

Lesser, J. (2013), *Immigration, ethnicity, and national identity in Brazil, 1808 to the present*, Cambridge University Press.

Lewis, B. & Kakkar, D. (2016), 'Harvard CGA Geotweet Archive v2.0'.
URL: https://doi.org/10.7910/DVN/3NCMB6

Luxen, D. & Vetter, C. (2011), Real-time routing with openstreetmap data, *in* 'Proceedings of the 19th ACM SIGSPATIAL international conference on advances in geographic information systems', pp. 513–516.

Marques, R. (2024), 'Forced migration and violent crime: Evidence from the venezuelan exodus to brazil', *The Journal of Development Studies* pp. 1–23.

Martén, L., Hainmueller, J. & Hangartner, D. (2019), 'Ethnic networks can foster the economic integration of refugees', *Proceedings of the National Academy of Sciences* **116**(33), 16280–16285.

Maurel, M. & Tuccio, M. (2016), 'Climate instability, urbanisation and international migration', *The Journal of Development Studies* **52**(5), 735–752.

Mazzoli, M., Diechtiareff, B., Tugores, A., Wives, W., Adler, N., Colet, P. & Ramasco, J. J. (2020), 'Migrant mobility flows characterized with digital data', *PloS one* **15**(3), e0230264.

McGregor, J. (1994), 'Climate change and involuntary migration: implications for food security', *Food policy* **19**(2), 120–132.

Moore, W. H. & Shellman, S. M. (2007), 'Whither will they go? a global study of refugees' destinations, 1965–1995', *International Studies Quarterly* **51**(4), 811–834.

Morales-Arilla, J. (2024), Autocrats in crisis mode: Strategic favoritism during economic shocks. Working Paper.

Morales-Arilla, J. & Gadgin Matha, S. (2024), 'Glocal: A global development dataset of subnational administrative areas', *Scientific Data* **11**(1), 851.

Mossaad, N., Ferwerda, J., Lawrence, D., Weinstein, J. & Hainmueller, J. (2020), 'In search of opportunity and community: Internal migration of refugees in the united states', *Science advances* **6**(32), eabb0295.

Natural Earth (2024), '10m coastline'. Accessed: 2024-10-02.
URL: https://www.naturalearthdata.com/downloads/10m-physical-vectors/10m-coastline/

Obolensky, M., Tabellini, M. & Taylor, C. (2024), Homeward bound: How migrants seek out familiar climates, Technical report, National Bureau of Economic Research.

Olivieri, S., Ortega, F., Rivadeneira, A. & Carranza, E. (2022), 'The labour market effects of venezuelan migration in ecuador', *The Journal of Development Studies* **58**(4), 713–729.

Olteanu, A., Castillo, C., Diaz, F. & Kıcıman, E. (2019), 'Social data: Biases, methodological pitfalls, and ethical boundaries', *Frontiers in big data* **2**, 13.

Petersen, W. (1958), 'A general typology of migration', *American sociological review* **23**(3), 256–266.

Piguet, E. (2013), 'From "primitive migration" to "climate refugees": The curious fate of the natural environment in migration studies', *Annals of the Association of American Geographers* **103**(1), 148–162.

Pirovino, S. & Papyrakis, E. (2023), 'Understanding the global patterns of venezuelan migration: determinants of an expanding diaspora', *Development Studies Research* **10**(1), 2147561.

Ramos, R. (2016), 'Gravity models: a tool for migration analysis', *IZA World of Labor* .

Rikani, A. & Schewe, J. (2021), 'Global bilateral migration projections accounting for diasporas, transit and return flows, and poverty constraints.', *Demographic Research* **45**, COV4–COV4.

Rozo, S. V. & Vargas, J. F. (2021), 'Brothers or invaders? how crisis-driven migrants shape voting behavior', *Journal of Development Economics* **150**, 102636.

Ryu, H. & Paudel, J. (2022), 'Refugee inflow and labor market outcomes in brazil: Evidence from the venezuelan exodus', *Population and Development Review* **48**(1), 75–96.

Saa, I. L., Novak, M., Morales, A. J. & Pentland, A. (2020), 'Looking for a better future: modeling migrant mobility', *Applied Network Science* **5**(1), 70.

Spyratos, S., Vespe, M., Natale, F., Weber, I., Zagheni, E. & Rango, M. (2018), 'Migration data using social media: a european perspective'.

Sviatschi, M., Cabra, N. & Rozo, S. (2024), Mass migration, the perpetuation of autocratic leaders and development in origin countries: Evidence from venezuela. Working Paper.

The World (2024), 'Millions of venezuelans living abroad will be unable to vote in upcoming presidential election'. Accessed: 2024-10-02.
URL: https://theworld.org/stories/2024/07/17/millions-of-venezuelans-living-abroad-will-be-unable-to-vote-in-upcoming-presidential-election

UNHCR (2023), 'Refugee statistics'. Accessed on 31 August 2024.
URL: https://www.unhcr.org/refugee-statistics/download/

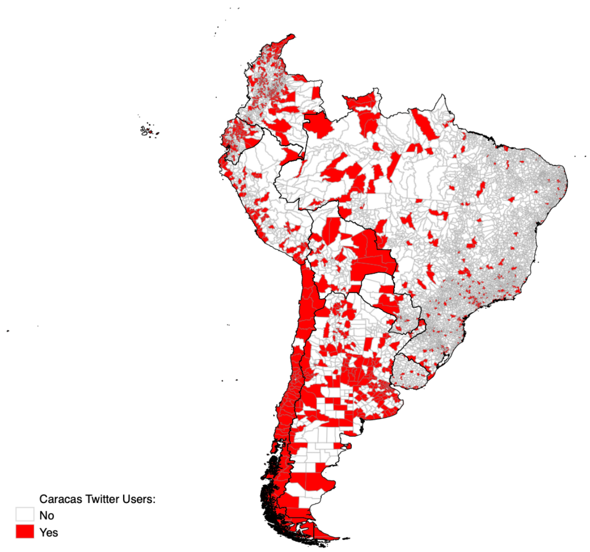
World Bank (2018), 'Migración desde venezuela a colombia: impactos y estrategia de respuesta en el corto y mediano plazo'.

Zagheni, E., Garimella, V. R. K., Weber, I. & State, B. (2014), Inferring international and internal migration patterns from twitter data, *in* 'Proceedings of the 23rd international conference on world wide web', pp. 439–444.

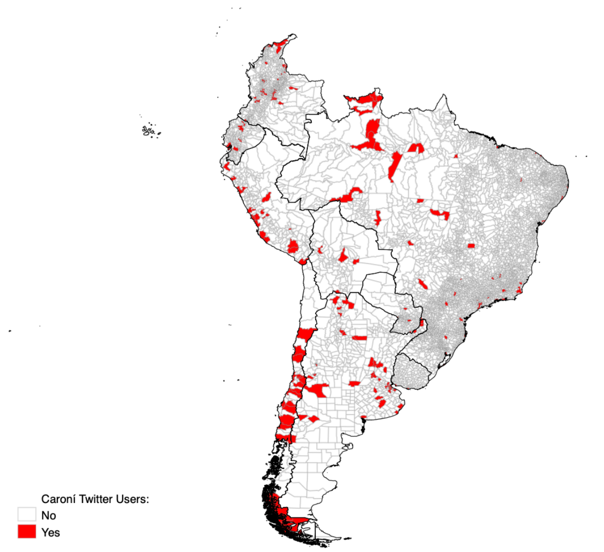
Zavodny, M. (1999), 'Determinants of recent immigrants' locational choices', *International Migration Review* **33**(4), 1014–1030.

Zorlu, A. & Mulder, C. H. (2008), 'Initial and subsequent location choices of immigrants to the netherlands', *Regional Studies* **42**(2), 245–264.

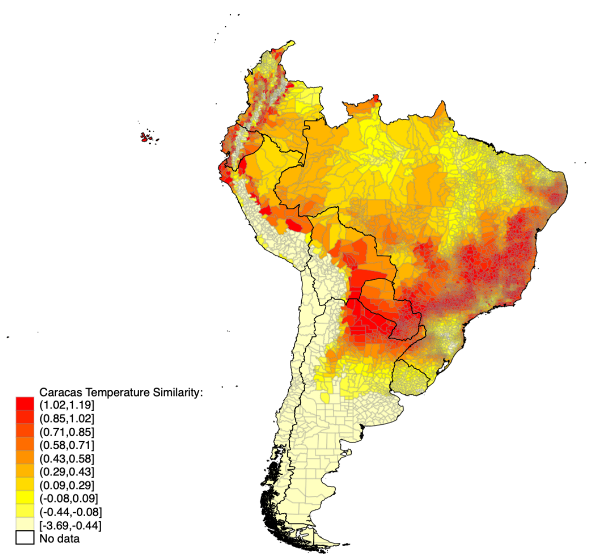
Figure 1: Twitter Flows and Temperature Similarities



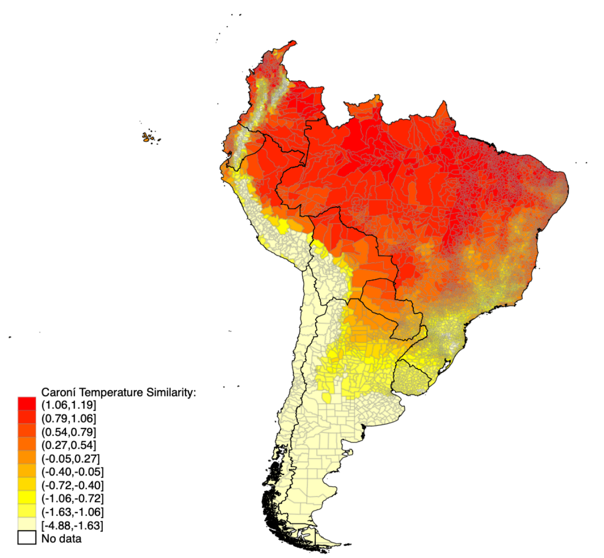
(a) Presence of Caracas Twitter Users



(b) Presence of Caroní Twitter Users



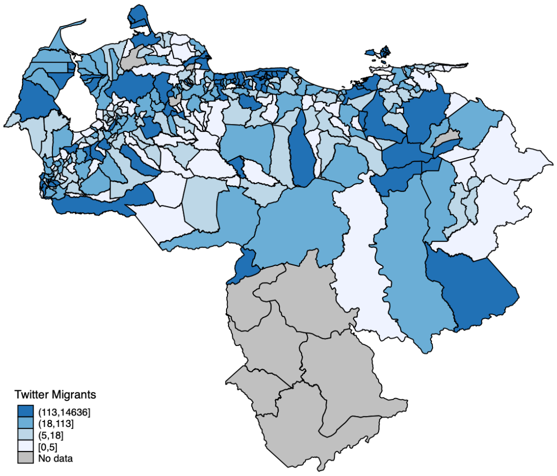
(c) Caracas Temperature Similarity



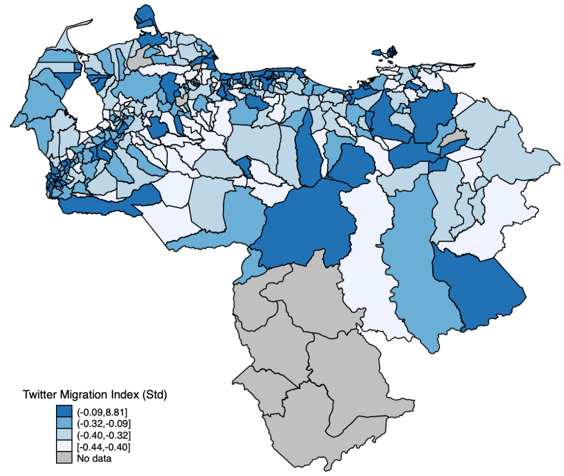
(d) Caroní Temperature Similarity

Note: Data derived from social media analysis using geolocated Twitter data, processed between 2011 and 2021. Ecological similarities are based on the GLocal dataset.

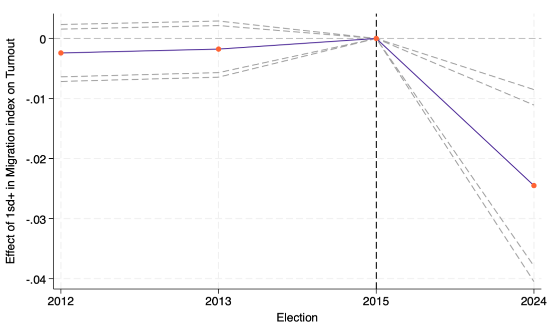
Figure 2: Effects of migration on 2024 electoral outcomes



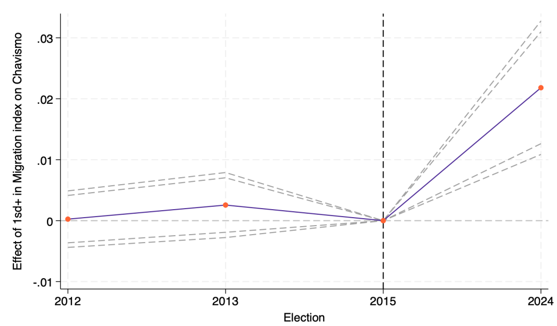
(a) Twitter flows



(b) Migration Index



(c) Migration Index → Turnout



(d) Migration Index → Regime support

Note: This figure illustrates the effects of migration on 2024 electoral outcomes, focusing on Twitter flows and the Migration Index, along with their impact on voter turnout and regime support.

Tables

Table 1: Validation analyses for Twitter-flows migration matrix

VARIABLES	Origin/Destination Level: *Pulso de la Migración* (IHS)		Destination Level: *Migración Colombia* (IHS)	
Twitter migrants (IHS)	0.625***	0.465***	0.852***	0.324***
	(0.0469)	(0.0471)	(0.0588)	(0.0350)
Total population (IHS)		-0.620***		0.960***
		(0.0428)		(0.0688)
Constant	0.150***	7.810***	4.625***	-4.643***
	(0.00662)	(0.530)	(0.191)	(0.693)
Observations	54,168	54,168	1,062	1,062
R-squared	0.244	0.258	0.382	0.630
Fixed Effects	Orig + Dest	Orig + Dest	None	State

Notes. Columns 1 and 2 follow the specification described in Equation 1, with standard errors clustered within each bilateral State combination. Columns 3 and 4 follow the specification described in Equation 2, with standard errors clustered at the State level. *** p<0.01, ** p<0.05, * p<0.1

Table 2: Gravity models of migration and ecological similarities of origins and destinations

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	Number of Twitter users from an origin in a destination							
Rainfall similarity	0.135***				0.118***	0.152***	0.124***	0.160***
	(0.0403)				(0.0375)	(0.0368)	(0.0380)	(0.0372)
Distance to coastline similarity		0.438***			0.388***	0.396***	0.355***	0.357***
		(0.113)			(0.108)	(0.100)	(0.108)	(0.0979)
Elevation similarity			0.00829		-0.0232	-0.0606***	-0.0171	-0.0538***
			(0.0221)		(0.0194)	(0.0184)	(0.0194)	(0.0175)
Temperature similarity				0.155**	0.184***	0.210***	0.177***	0.202***
				(0.0698)	(0.0681)	(0.0654)	(0.0684)	(0.0651)
Driving proximity						0.777***		0.791***
						(0.104)		(0.103)
Population density similarity							0.0173***	0.0209***
							(0.00514)	(0.00588)
Constant	2.365***	2.165***	2.377***	2.544***	2.373***	2.021***	2.440***	2.095***
	(0.0185)	(0.0489)	(0.0224)	(0.0785)	(0.0962)	(0.107)	(0.0931)	(0.102)
Observations	587,494	587,494	587,494	587,494	587,494	587,494	587,494	587,494
Fixed Effects	Orig + Dest	Orig + Dest	Orig + Dest	Orig + Dest	Orig + Dest	Orig + Dest	Orig + Dest	Orig + Dest

Notes. All columns follow the specification described in Equation 4, with standard errors clustered within each bilateral State combination. *** p<0.01, ** p<0.05, * p<0.1

Table 3: Gravity model predictions and Migration flows

VARIABLES	(1) Origin/Destination Level: *Pulso de la Migración* (IHS)	(2)	(3) Destination Level: *Migración Colombia* (IHS)	(4)
Predicted flows (IHS)	0.235***	0.200***	1.083	1.166***
	(0.0469)	(0.0440)	(0.653)	(0.306)
Total population (IHS)		-0.824***		1.254***
		(0.0491)		(0.0596)
Constant	-0.554***	9.719***	-3.935	-17.65***
	(0.148)	(0.585)	(5.844)	(2.833)
Observations	54,168	54,168	1,062	1,062
R-squared	0.217	0.245	0.053	0.608
Fixed Effects	Orig + Dest	Orig + Dest	None	Dept

Notes. Columns 1 and 2 follow the specification described in Equation 1, considering bilateral predictions from the gravity model outlined in Equation 4 on the right hand side. Standard errors clustered within each bilateral State combination. Columns 3 and 4 follow the specification described in Equation 2, considering bilateral predictions from the gravity model outlined in Equation 4 aggregated at the destination level on the right hand side. Standard errors clustered at the State level. *** p<0.01, ** p<0.05, * p<0.1

Online Appendix

A.1 Additional tables

Table A.1: Summary Statistics

Panel A: Origin/Destination Level

	N	mean	sd	min	max
Twitter Flows	2,724,618	0.0235	1.913	0	1,532
Pulso Flows	54,168	28.50	490.3	0	74,208
Predicted flows	2,722,590	11.59	4.166	1.195	23.68
Elevation similarity (Std)	2,723,266	0	1	-5.285	0.901
Rainfall similarity (Std)	2,723,942	0	1	-13.10	1.102
Temperature similarity (Std)	2,723,266	0	1	-5.318	1.190
Distance to coastline similarity (Std)	2,724,618	0	1	-4.615	1.010
Driving distance (Std)	2,724,618	0	1	-2.697	2.709
Population density similarity (Std)	2,724,618	0	1	-16.25	0.374
Bilateral population	2,724,618	115,342	267,470	10.20	13,128,492

Panel B: Destination Level

	N	mean	sd	min	max
Twitter Flows	8,061	7.944	128.0	0	5,222
Predicted Flows	8,061	3,915	1,208	0	5,866
Venezuelans (*Migración Colombia*)	1,062	2,308	18,417	0	495,236
Population	8,061	42,802	219,198	0	11,313,501
Elevation	7,949	665.5	767.3	1	4,569
Rainfall	8,038	3.293	2.974	0.529	12.10
Temperature	8,038	22.61	4.376	4.243	29.29
Distance to coastline	8,055	288.2	255.4	0.0857	1,360

Panel C: Origin Level

	N	mean	sd	min	max
Twitter Flows	335	189.6	882.1	0	14,636
Population	335	73,161	154,066	1,709	1,814,991
Elevation	335	564.6	624.9	1	3,266
Rainfall	335	5.120	0.689	4.034	10.74
Temperature	335	25.32	2.649	15.14	29.31
Distance to coastline	335	86.39	107.0	0.377	808.0
Migration Index	324	0	1	-0.44	8.814
2024 Electors	324	54761.71	98859.68	2729	1210692
Electoral Turnout	1,296	0.732	0.087	0.397	0.873
Regime Support	1,296	0.493	0.156	0.087	0.936

Notes. Panel A shows summary statistics for variables at the bilateral level, defined by every combination between Venezuelan second level administrative units and those of other South American countries. Panel B shows summary statistics for variables at the destination level, defined by second level administrative units in South America outside of Venezuela. Panel B shows summary statistics for variables at the origin level, defined by second level administrative units in Venezuela.

Table A.2: Gravity Models: Flows between cities + Flows outside capitals

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	Number of Twitter users from an origin in a destination					
Rainfall similarity (Std)	0.118***	0.129	0.121***	0.160***	0.164*	0.162***
	(0.0375)	(0.0869)	(0.0376)	(0.0372)	(0.0874)	(0.0371)
Distance to coastline similarity (Std)	0.388***	0.342***	0.409***	0.357***	0.246**	0.378***
	(0.108)	(0.124)	(0.109)	(0.0979)	(0.111)	(0.0978)
Elevation similarity (Std)	-0.0232	-0.0344	-0.0290	-0.0538***	-0.0593*	-0.0625***
	(0.0194)	(0.0369)	(0.0192)	(0.0175)	(0.0342)	(0.0172)
Temperature similarity (Std)	0.184***	0.190	0.187***	0.202***	0.206	0.208***
	(0.0681)	(0.192)	(0.0678)	(0.0651)	(0.186)	(0.0648)
Constant	2.373***	3.900***	2.111***	2.095***	3.902***	1.831***
	(0.0962)	(0.250)	(0.0957)	(0.102)	(0.216)	(0.102)
Observations	587,494	16,920	587,485	587,494	16,920	587,485
Fixed Effects	O + D	O + D	O + D	O + D	O + D	O + D
Distance/Density Controls	No	No	No	Yes	Yes	Yes
Sample	All	Non-Rural Pairs	Non-Capital Pairs	All	Non-Rural Pairs	Non-Capital Pairs

Notes. All columns follow the specification described in Equation 4, with standard errors clustered within each bilateral State combination. *** p<0.01, ** p<0.05, * p<0.1

Table A.3: Gravity Models: Flows outside Venezuelan borders

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	Number of Twitter users from an origin in a destination					
Rainfall similarity (Std)	0.118***	0.0496	0.0649**	0.160***	0.0840**	0.0977***
	(0.0375)	(0.0414)	(0.0263)	(0.0372)	(0.0404)	(0.0258)
Distance to coastline similarity (Std)	0.388***	0.216***	0.233**	0.357***	0.172**	0.208***
	(0.108)	(0.0815)	(0.0912)	(0.0979)	(0.0727)	(0.0802)
Elevation similarity (Std)	-0.0232	-0.0214	-0.0250	-0.0538***	-0.0507***	-0.0482***
	(0.0194)	(0.0189)	(0.0183)	(0.0175)	(0.0170)	(0.0160)
Temperature similarity (Std)	0.184***	0.214***	0.178***	0.202***	0.235***	0.190***
	(0.0681)	(0.0656)	(0.0646)	(0.0651)	(0.0632)	(0.0617)
Constant	2.373***	2.604***	2.560***	2.095***	2.407***	2.436***
	(0.0962)	(0.0879)	(0.0862)	(0.102)	(0.0878)	(0.0827)
Observations	587,494	527,001	538,200	587,494	527,001	538,200
Fixed Effects	O + D	O + D	O + D	O + D	O + D	O + D
Distance/Density Controls	No	No	No	Yes	Yes	Yes
Sample	All	No Vzla Border	No Col-Bra Border	All	No Vzla Border	No Col-Bra Border

Notes. All columns follow the specification described in Equation 4, with standard errors clustered within each bilateral State combination. *** p<0.01, ** p<0.05, * p<0.1