Escuela de Gobierno y Transformación Pública
Tecnológico de Monterrey

Industrial Matching in Refugees' Settlement Choices: Evidence from Venezuelans in Colombia

José Morales-Arilla
josemoralesarilla@tec.mx
Escuela de Gobierno y Transformación Pública, Tecnológico de Monterrey

Rui Alleyne
ralleyne@ucab.edu.ve
Universidad Católica Andrés Bello, Venezuela

Omar Hammoud Gallego
omar.hammoud-gallego@durham.ac.uk
Durham University, UK

School of Government and Public Transformation

Working Paper No. 38

Publication update: August 2026

Industrial Matching in Refugees' Settlement Choices: Evidence from Venezuelans in Colombia.

Jose Morales-Arilla* Rui Alleyne† Omar Hammoud Gallego‡

August 10, 2026

Abstract

This study provides evidence favoring the "industrial matching" hypothesis that refugees sort between areas with similar industries where home-grown skills are transferable. We evaluate flows between Venezuelan and Colombian localities in the context of Venezuela's refugee shock. We perform Poisson pseudo-maximum likelihood (PPML) gravity models to evaluate whether bilateral similarities in baseline industrial structure explain refugee flows after controlling for key potential confounders. Venezuelan refugees tend to move towards Colombian localities with similar baseline industries relative to their origin localities. Heterogeneity analyses highlight agriculture as a particularly relevant sector behind our results. Despite limited agency, refugees may sort towards economic opportunities consistent with their background - a relevant finding for policymakers concerned about the economic integration of the refugees they host.

Keywords: Refugees, Industrial Matching, Venezuela, Colombia.
JEL: J61, R23, J15, O15

*Research Professor at Tecnológico de Monterrey, Mexico. josemoralesarilla@tec.mx
†Adjunct Professor at Universidad Católica Andrés Bello, Venezuela. ralleyne@ucab.edu.ve
‡Assistant Professor in Public Policy, Durham University, UK. omar.hammoud-gallego@durham.ac.uk

Introduction

According to the United Nations' High Commissioner for Refugees, in 2025 there were over 129 Million forcibly displaced people across the globe, most of whom are hosted in low and middle income countries (UNHCR, 2026). Yet, while extensive research has established the factors driving refugees away from their countries, such as the presence of conflict, insecurity, economic wealth and opportunities (Van Hear, Bakewell and Long, 2018; Adhikari, 2013; Matsui and Raymer, 2020; Conte and Migali, 2019; Kang, 2021), there is still limited scholarship about which factors determine where refugees decide to settle having initially left their country of origin. The aim of our study is to evaluate whether, even in contexts of forced displacement, refugees move to areas that have a similar industrial structure to their locations of origin. This "industrial matching" hypothesis would extend a classic insight from the economics of migration—that migrants locate where the returns to their skills are highest (Borjas, 2001), and that occupational skill endowments shape migration decisions (Patt et al., 2021)—to a setting of forced displacement, where such deliberate economic sorting is far from guaranteed.

To test the hypothesis of "industrial matching" in refugee settlement choices, we focus on the case of the sudden massive displacement of Venezuelan refugees in Colombia.[1] We leverage statistical and social-network sources capturing the local origin-destination migration flows from Venezuela into Colombia. Moreover, we consider baseline census data from both countries to measure the local employment in every industry, and then calculate the correlation in origin-destination sectoral structures as our measures of bilateral industrial similarity. Using Poisson Pseudo-Maximum Likelihood (PPML) gravity models of bilateral flows (Correia, Guimarães and Zylkin, 2020) controlling for key confounding factors,[2] we find

[1] While Venezuelans are only recognised as refugees in a handful of countries, mainly Brazil and Mexico, scholars and international organisations have long argued that the conditions in Venezuela mean they do meet that definition (R4V, 2026; Freier, Berganza and Blouin, 2022; Blouin, Berganza and Freier, 2020; Hammoud-Gallego and Freier, 2023).

[2] We consider controls for bilateral distance, ecological similarity, population density similarity and the presence of bilateral social-networks.

that refugees are more likely to move to municipalities in Colombia that are most similar in terms of their economic structure to those of their Venezuelan origins. Chiefly, the effects of industrial similarity on bilateral refugee flows are about 6% those of geographic proximity and about 70% those of social network intensity, two key determinants of bilateral migration flows established in the literature.

Our results are robust to a wide battery of robustness tests considering alternative geographic and industrial aggregation levels, construction of the treatment variable, specific control variables, and outcome variable sources. Finally, in an attempt to identify what industries drive our results, we evaluate whether iteratively removing certain sectors of the economy from the construction of our treatment variable attenuates the estimated effects. We find that removing agriculture from the analysis reduces the estimated effect size to the point of compromising statistical significance, and this pattern holds regardless of the inclusion of population density controls. Taken together, the evidence suggests that Venezuelan refugees in Colombia settle in locations that are industrially similar to their Venezuelan origins, and that agriculture plays an important role behind this pattern.

Existing scholarship has tried to identify the determinants of economic migrants' and refugees' destination choices. These include a variety of factors, such as real wages and moving costs (Kennan and Walker, 2011), geographical, cultural and ethnic closeness (Fafchamps and Shilpi, 2013; Wang, De Graaff and Nijkamp, 2016; Tubadji and Nijkamp, 2015), former colonial ties (Matsui and Raymer, 2020), climate similarity (Morales-Arilla, Martin and Morales, 2025; Obolensky, Tabellini and Taylor, 2024; Lesser, 2013), and pre-existing networks (Munshi, 2003; Stuart and Taylor, 2021), among others (Spörlein, 2015; de Haas et al., 2019). However, to our knowledge, little research has been dedicated to testing the possible role of industrial similarities between the areas of origin and destination as a predictive factor for migratory choices in contexts of forced displacement and limited settlement agency. Specifically, *we hypothesize that refugees will disproportionately relocate to municipalities in the host country that are similar in their employment structure to their area of origin.* At

the margin, refugees may be more likely to move to areas where employment structures –the type of industries people work in– are most similar to those in their areas of origin, as this means potentially higher demand for the skills and know-how acquired on the job back home. Indeed, Bazzi et al. (2016) document important economic benefits of skill transferability for internal Indonesian migrants randomly assigned across agroclimatically similar villages in different islands. Our findings are consistent with refugees settling in locations where the benefits of skill transferability are available to them.

Whether refugees end up in locations where their skills are in demand is also the central concern of a growing literature on refugee placement. Studies exploiting the quasi-random assignment of refugees across localities under governmental dispersal policies show that initial placement conditions—particularly local labor market conditions at arrival—have lasting effects on refugees' economic integration (Edin, Fredriksson and Åslund, 2003; Åslund and Rooth, 2007; Damm, 2009; Aksoy, Poutvaara and Schikora, 2023). Building on this evidence, recent work develops algorithmic assignment mechanisms that place refugees in the locations where their expected integration outcomes are highest (Bansak et al., 2018; Ahani et al., 2021; Acharya, Bansak and Hainmueller, 2022). These literatures study contexts in which settlement locations are assigned by host governments. Much less is known about the reverse scenario: whether refugees who retain agency over their settlement choices sort themselves toward locations that demand the skills they bring from home.

This study contributes to the literature on forced displacement and refugee settlement choices. To our knowledge, we are the first to assess the relevance of industrial matching in migrants' settlement choices in a context of forced displacement across international borders. Our findings suggest that refugees themselves search out the economic benefits of skill transferability, organically approximating the type of skill–location matching that assignment mechanisms are designed to engineer. The case we study—Venezuelan forced displacement into Colombia—is particularly well suited to this question for three reasons. First, the two countries share the same language, religion, ethnicity, history and culture, as

well as a porous border over 2,000 kilometres long, allowing us to hold constant many of the main determinants of settlement choices established in the literature. Second, Venezuelans in Colombia retained an unusual degree of settlement agency: the Colombian government's limited capacity to manage the inflow, together with its decision to ease refugees' legal regularization, meant that arriving Venezuelans were neither dispersed across the territory nor restricted in their movement, so that observed settlement patterns largely reflect refugees' own choices. Finally, the setting is important in its own right: the Venezuelan displacement crisis is the biggest displacement event in the Western hemisphere in modern times, with an estimated 8 million people having left the country since 2016 due to dire political, social and economic conditions at home, over 2.9 million of whom moved to neighbouring Colombia (R4V, 2026; Hammoud-Gallego, 2024).

The paper proceeds as follows. Section I introduces the empirical setting. Section II presents our data. Section III describes our research strategy. Section IV presents our results, and Section V concludes with a discussion on the implications of our results.

1 Empirical Setting: The Venezuelan Refugee Shock

Massive displacement from Venezuela is a relatively recent phenomenon (Freier and Parent, 2019). Historically, the country had been a recipient of migrants and refugees, especially from neighbouring Colombia, due to the latter's high levels of violence during its civil conflict (Gottwald, 2004). Yet, due to Venezuela's slide into authoritarianism under the presidencies of Hugo Chávez first, then Nicolás Maduro, coupled with the economic collapse of the country - with a 75% decrease in GDP between 2014 and 2021 - this migratory trend reversed. Current estimates are that 8 million Venezuelans (over 25% of the population) have left the country since 2016, 7 million of which moved to other Latin American countries. With 2.9 million migrants, Colombia was the main recipient of the Venezuelan refugee shock (R4V, 2026; on Foreign Relations, 2024; Corrales, 2023). Figure A.1 shows the total number of

Venezuelan migrants in Colombia between 2014 and 2024, capturing the absence of migrants before 2016.

Extensive research has been dedicated to the study of Venezuelan displacement in Colombia and throughout the region (Hammoud-Gallego, 2024), with some studies focusing on the effects of the mass Venezuelan regularization programs implemented by the Colombian government (Ibáñez et al., 2025; Bahar, Ibáñez and Rozo, 2021; Ibáñez et al., 2026). Similarly, studies have considered the effects of the Venezuelan migration on crime (Knight and Tribin, 2023; Marques, 2025), as well as migrants' experiences in host countries (Freier and Pérez, 2021). Yet, with the exception of Morales-Arilla, Martin and Morales (2025), no research has tried to understand how forcibly displaced Venezuelans from different parts of the country have sorted themselves within specific host nations.

2 Data Sources and Key Measures

We measure local origin-destination migration flows between Venezuela and Colombia using three complementary data sources. First, we use the VenRePS Survey, a longitudinal survey of Venezuelan forced migrants in Colombia. The survey contains information on migrants' place of origin in Venezuela and current location in Colombia, which allows us to construct bilateral migration flows between Venezuelan and Colombian municipalities (Ibáñez, Moya and Rozo, 2021). Second, we use Colombia's *Pulso de la Migración* (Migration Pulse) survey data, a series of representative cross-sectional surveys, which also contains information on migrants' origins and destinations across locations in both countries (DANE, 2025). Finally, we construct a Twitter/X-based migration measure using geolocated Twitter activity (Lewis and Jain, 2016). Specifically, we identify users tweeting from Colombia and recover the Venezuelan municipalities from which these users previously tweeted, allowing us to infer municipality-to-municipality migration links using digital traces (Morales-Arilla, Martin and Morales, 2025). We collect the count of bilateral flows for each origin/destination pair from

these three sources, considering the VenRePS count as our main outcome variable given its broader scope capturing Venezuelan migrants in Colombia.

To measure economic similarity across bilateral pairs, we construct employment structure similarity indices using census and household survey data from Venezuela and Colombia at both the municipality and the state level. At the municipality level, we use the 2001 Venezuelan Population Census from Venezuela's *Instituto Nacional de Estadística* (INE) and the 2005 Colombian Population Census from Colombia's *Departamento Administrativo Nacional de Estadística* (DANE).[3] These censuses allow us to construct comparable labor market industry structures for 335 Venezuelan municipalities and 529 Colombian municipality aggregations prior to the onset of the migratory crisis after 2016. Although these datasets provide substantial geographic detail, the industrial classifications are relatively aggregated. After harmonizing industry classifications across countries, we are able to consistently identify 15 broad macro-industries common to both censuses. At the state–department level, we use labor force surveys with substantially more detailed industrial information. For Venezuela, we use INE's 2014 Venezuelan *Encuesta de Hogares por Muestreo*, while for Colombia we use DANE's 2015 *Gran Encuesta Integrada de Hogares (GEIH).* These datasets cover all Venezuelan states, but only 24 out of the 32 Colombian departments (equivalent to states). Although the geographic aggregation is coarser than possible with the census data, these surveys contain much more detailed information on economic activities, allowing us to harmonize a total of 73 industries across both countries.

For all sources, we collect a measure of the number of employees $E_{i,l}$ working on a specific industry i in location l. With this, we can calculate the local share of employees in a specific industry ($S_{i,l} = E_{i,l} / \sum_i E_{i,l}$), and importantly, the location quotient of the industry in the location. The location quotient of industry i in location l captures the relevance of the industry as an employer in the location relative to the importance of the industry for the whole country. It is calculated as follows:

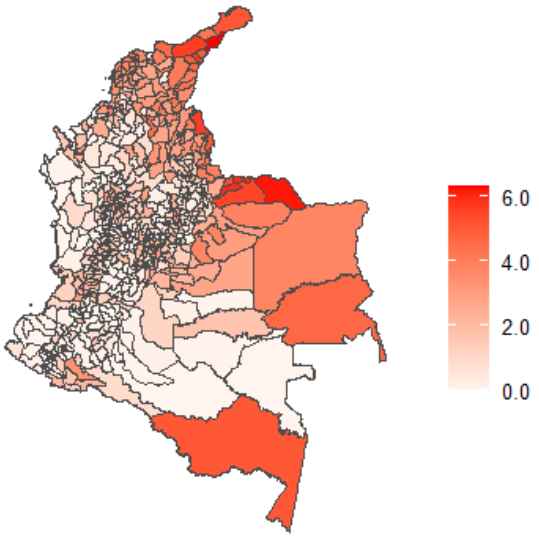
[3]We would like to thank Colombia's DANE for access to data for the 2005 Colombian Census and 2015 *Gran Encuesta de Hogares* survey data.

$$LQ_{i,l} = \frac{E_{i,l}/\sum_i E_{i,l}}{\sum_l E_{i,l}/\sum_i \sum_l E_{i,l}} = \frac{S_{i,l}}{S_i} \tag{1}$$

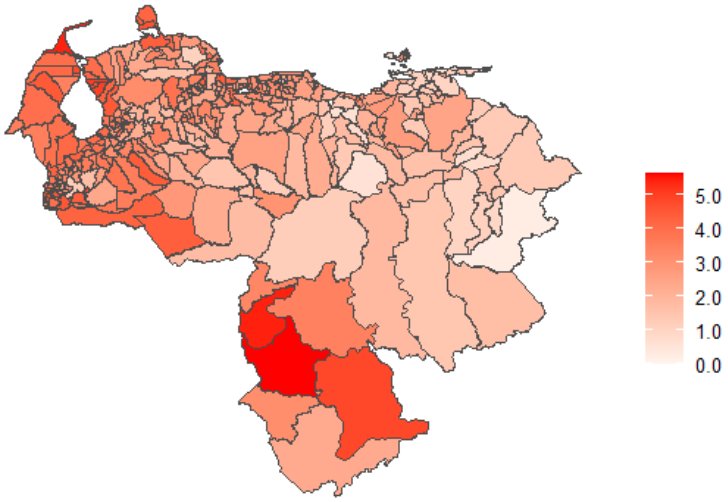
Having calculated $LQ_{i,l}$ and $S_{i,l}$, we can now measure the correlation across industries of these scores for every pair of municipalities (or states) of Venezuelan origin and of Colombian destination. These correlations are our measures of baseline industrial similarity between origins and destinations. Because some industries tend to be relatively big (e.g. commerce) or small (e.g. manufacture) across different types of cities, we consider location quotient based measures of industrial similarity for our main specification. Moreover, the fact that drastically different labor markets often get grouped within the same states could add noise to our bilateral similarity metrics. For this reason, we choose municipality-level bilateral analyses for our main specifications.

Finally, we collect bilateral driving distances; bilateral ecological similarities in terms of rainfall, distance to coast, elevation and temperature; and population density similarities from Morales-Arilla, Martin and Morales (2025). Moreover, we collect bilateral, local measures of social connectedness on Facebook from Johnston et al. (2026). These variables will provide controls for first order potential confounders in our analyses: geographic proximity, ecological and population density similarity, and the presence of bilateral social networks. All treatment and control variables in our analysis are standardized for comparability of coefficients and average marginal effects. Tables A.1 and A.2 provide summary statistics of our data at the bilateral municipality and state level, respectively. Figure 1 shows a measure of the total flows to Colombian municipalities (Panel A) and from Venezuelan municipalities (Panel B) according to the VenRePS data. Figure 2 provides measures of the flows from the VenRePS data and LQ similarity scores across Colombian Municipalities from two Venezuelan origin municipalities: Caracas and Caroni. Figure A.2 repeats this analysis across Colombian departments with a a more detailed industrial classification for two Venezuelan states of origin: Caracas and Bolívar.

Figure 1: VenRePS Migration Flows by Destination and by Origin



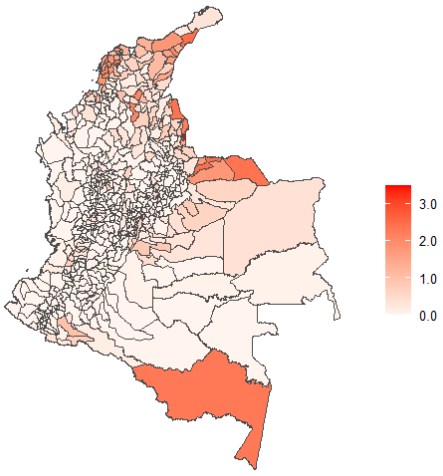
(a) Migrants received by Colombian municipality



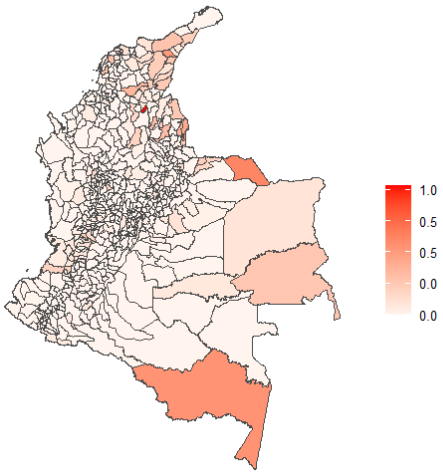
(b) Migrants from Venezuelan municipalities

Notes: VenRep flows correspond to the inverse hyperbolic sine transformation (`asinh`) of the number of VenRep flows per 1,000 inhabitants in each Colombian municipality.

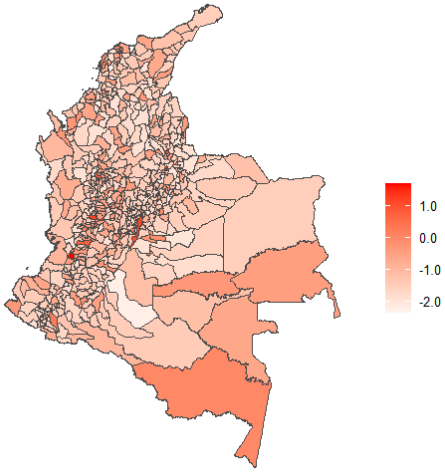
Figure 2: VenRep Flows and Employment Similarities



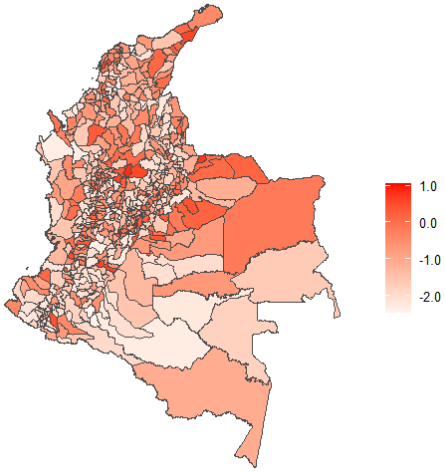
(a) Caracas migrants presence (VenReps)



(b) Caroni migrants presence (VenReps)



(c) Caracas Employment Similarity



(d) Caroni Employment Similarity

Notes: VenRep flows correspond to the inverse hyperbolic sine transformation (`asinh`) of the number of VenRep flows per 1,000 inhabitants in each Colombian municipality. Employment similarities correspond to the standardized location quotient (LQ) similarity measure across municipalities.

3 Empirical Strategy

We examine whether Venezuelan refugees disproportionately relocate to Colombian locations with similar economic structures. To do so, we estimate bilateral migration flows between origin locations in Venezuela and destination locations in Colombia using Poisson pseudo-maximum likelihood (PPML) models with high-dimensional fixed effects. Our baseline specification is:

$$Y_{od} = \exp\left(\beta \times \text{Similarity}_{od} + \lambda_o + \lambda_d\right) + \varepsilon_{od} \tag{2}$$

where Y_{od} represents migration flows or migrant presence between Venezuelan origin location o and Colombian destination location d. Depending on the specification, the outcome variable corresponds to: (i) migrant observations from the VenRePS Survey, (ii) migrant counts from Colombia's *Pulso de la Migración* survey, or (iii) the number of Twitter/X users inferred to have migrated from Venezuelan locations to Colombian destinations using geolocated digital traces.The main explanatory variable, $Similarity_{od}$, measures the degree of economic similarity between Venezuelan and Colombian locations. We construct two alternative measures. The first is a revealed comparative advantage (LQ) similarity index based on the correlation of standardized LQ vectors across industries. The second is an employment structure similarity index based on the correlation of industry employment shares (S) between locations. λ_o and λ_d represent origin and destination fixed effects that absorb all time-invariant characteristics specific to origins and destinations, including local economic size, historical migration intensity, and destination attractiveness. ε_{od} capture the error term. Standard errors are clustered allowing for within origin and within destination correlation in errors.

We add a number of bilateral controls to this baseline specification. First, we control for standardized driving distance between origin and destination locations, which captures migration costs and geographic frictions. Second, we include ecological similarity controls

constructed from absolute differences in rainfall, elevation, temperature, and distance to the coast. These controls account for the possibility that migrants sort across locations with similar environmental conditions. In later specifications, we add two relevant but potentially problematic control variables. First, we control for population density similarities to capture urbanization and settlement patterns.[4] Second, we consider the intensity of bilateral social connections between localities in Venezuela and Colombia through Facebook's Social Connectedness Index (SCI) (Johnston et al., 2026).[5]

In addition to reporting coefficient estimates, we compute average marginal effects (AMEs) for both industrial similarity and our measures of geographic proximity and social connectedness - two key determinants of bilateral migration well established in the literature. We also report the ratio between the AME of industrial similarity and that of geographic distance and social connectedness to benchmark the relative importance of industrial matching compared to that of geographic proximity and of the existence of social networks between origins and destinations in shaping migrant location choices. This will help us provide a sense of the economic significance of our estimates. While robust to alternative specifications on all dimensions, our preferred specification takes bilateral flows at the municipality level from the VenRePS dataset, measures industrial similarity through the correlation of LQ-based measures of industry presence in a location, and avoids controls for potential bad controls.

The identifying assumption necessary to interpret our estimates as causal effects is for potential migration flows to be independent of levels of industrial similarity conditional

[4]While we show that our main results are robust to controlling for population density similarities, such controls are not part of our preferred specification. It is possible that the agroindustrial potential of a labor market shapes its demographic density. If this was the case, population density similarities could work as a mediator between industrial similarity and bilateral migration, constituting a bad control.

[5]Johnston et al. (2026) describe the SCI as a highly stable object capturing the geographic structure of real-world social networks. However, the published SCI measure is calculated on a network of Facebook friendships from January of 2026. The timing of this measurement opens the possibility that SCI values are endogenous to the refugee shock. If this was the case, the SCI would constitute another bad control in our setting, as it may be a consequence of our outcome variable and may work as a collider between our outcome and treatment variables. For this reason, the SCI is not part of our preferred specification. Still, the bias that would be induced by this possibility would lead us to understate the true positive effect of industrial similarity on bilateral migration. Hence, at worst, controlling for the SCI would allow us to estimate a lower-bound of the true effect we hope to measure. As we show below, our results are robust to the inclusion of the SCI in our regression analyses.

on observed confounders. Since we do not leverage exogenous sources of variation in the assignment of bilateral industrial similarities, it is difficult to establish the plausibility such an assumption. Hence, one important limitation to our study is that our research design is correlational, and results should be considered as such. However, our selection of empirical setting and strategy helps us identify a relationship that is orthogonal to most first-order determinants of migrant settlement choices established in the literature. As mentioned above, our setting considers South-South displacement across culturally, ethnically and historically similar countries. The origin and destination fixed effects in our regression specifications absorb all pull factors shared across origins and all push factors shared across destinations. Finally, our analysis controls for most first-order determinants of bilateral migration flows: Geographic proximity, social network intensity and ecological and demographic similarities. Consequently, it is hard to identify plausible unobserved confounders that independently explain our robust and economically meaningful estimates without appealing to Venezuelan refugees' rational search for the economic returns that come with skill transferability.

4 Results

We test our industrial matching hypothesis by running a series of Poisson Pseudo-Maximum Likelihood gravity models on bilateral flows from Venezuela to Colombia. Based on our preferred specification (municipality level bilateral flows from the VenRePS dataset and LQ-based industrial similarities), Table 1 provides estimates of Equation 2 (Column 1) and progressively adds geographic proximity (Column 2) and ecological similarity variables (Column 3) as regression controls. Estimates of the β coefficient all yield positive and statistically significant results. The corresponding AME estimates suggest that an increase of one standard deviation in industrial similarity increases the expected bilateral flows by 0.24-0.30 refugees, representing between 4.3%-6.3% of the AME of geographic proximity.

Table 1: Venezuelan to Colombian Municipalities: Migration Flows and LQ Similarity

	(1)	(2)	(3)
Industrial Similarity - LQ (sd)	0.117***	0.114***	0.090**
	(0.040)	(0.041)	(0.040)
AME Industrial Similarity (sd)	0.310***	0.304***	0.240**
	(0.106)	(0.102)	(0.100)
AME Driving distance (sd)	–	7.018***	3.802***
	–	(1.589)	(1.238)
AME Ratio: Ind. Sim. / Distance	–	0.043	0.063
Observations	166,332	166,332	165,998
Origin Muni. FE	Yes	Yes	Yes
Dest. Muni. FE	Yes	Yes	Yes
Driving distance	No	Yes	Yes
Ecological controls	No	No	Yes

Notes: The table reports municipality-to-municipality PPML estimates of migration flows using the VenRePS Survey. The dependent variable is the number of migration observations, and the main explanatory variable is the standardized LQ correlation between municipalities. All specifications include origin and destination municipality fixed effects. Column (2) additionally controls for standardized driving distance, while column (3) also includes standardized absolute differences in ecological characteristics (rainfall, distance to coast, elevation, and temperature). Standard errors are reported in parentheses and are clustered at the Venezuelan state–Colombian department pair level, where clusters are defined by unique combinations of Venezuelan state and Colombian department. Average marginal effects (AMEs) are reported for LQ correlation and driving distance where applicable. The ratio is computed as the AME of LQ correlation divided by the AME of driving distance. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 2 extends these specifications by adding controls for bilateral population density similarity and social network connectedness.[6] Column 1 restates the results from Column 3 in Table 1, Column 2 controls for population density similarities, Column 3 controls for bilateral social connectedness, and Column 4 controls for both. Estimates of β remain positive and statistically significant, while the value of the AME suggests that an increase in 1 s.d. in industrial similarity increases the expected bilateral flows by 0.20-0.22. As Columns 3 and 4 control for social connectedness, we can compare the AMEs associated with industrial similarity to that of social network intensity, finding that the former captures between 67%-70% of the latter.

We perform a number of robustness checks iterating over the main aspects of our empirical framework, and report results in the appendix. Table A.3 replicates the analyses in Table 1 but considers employment share based (S-based) measures of industrial similarity. Table A.4 also replicates the analysis of Table 1 but works with state-level aggregates of bilateral flows, industrial similarities,[7] and control variables. Table A.5 replicates the result of Column 3 in Table 1, but iterates over the source of bilateral flows between VenRePS (Column 1), *Pulso de la Migración* (Column 2) and Twitter/X-based flows (Column 3). Finally, Table A.6 replicates the iteration over outcome sources from Table A.5, but working on state-level data.[8] Overall, conclusions regarding the statistical and economic significance of the effects of industrial similarity on bilateral migration flows remain unaffected, with the exception of Column 2 of Table A.5, which yields a smaller and non-statistically significant estimate for municipality-level specifications considering flows from *Pulso de la Migración*.

Finally, to assess whether our results are driven by specific macro-industries, we re-run our models with a leave-one-industry-out strategy in measuring the bilateral industrial

[6] As mentioned above, while they capture important potential confounders of the relationship between industrial similarity and bilateral migration, they may also work as either mediators or colliders in that relationship. In that event, conditioning for them may introduce bias into our estimates.

[7] Based on the more detailed industrial classification available in baseline state-level employment surveys from both countries.

[8] Importantly, because the estimated effects of geographic proximity cannot be precisely measured for state-level aggregates, we do not compare the AME of industrial similarity to that of geographic proximity in Tables A.4 and A.6.

Table 2: Migration Flows and LQ Similarity: Population Density and Facebook SCI controls

	(1)	(2)	(3)	(4)
Industrial Similarity - LQ (sd)	0.090**	0.084**	0.071**	0.069**
	(0.040)	(0.039)	(0.034)	(0.035)
AME Ind. Similarity (sd)	0.240**	0.224**	0.209**	0.201**
	(0.100)	(0.099)	(0.095)	(0.099)
AME Facebook SCI (sd)	–	–	0.299***	0.298***
	–	–	(0.040)	(0.040)
AME Ratio: Ind. Sim. / SCI	–	–	0.699	0.672
Observations	165,998	165,998	138,670	138,670
Origin Muni. FE	Yes	Yes	Yes	Yes
Dest. Muni. FE	Yes	Yes	Yes	Yes
Driving distance	Yes	Yes	Yes	Yes
Ecological controls	Yes	Yes	Yes	Yes
Population density	No	Yes	No	Yes
Facebook SCI	No	No	Yes	Yes

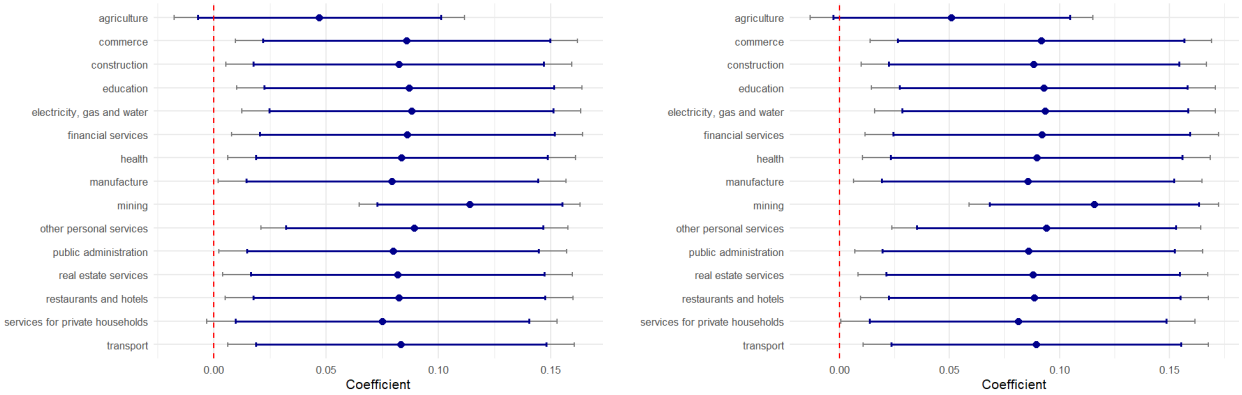
Notes: The table reports municipality-to-municipality PPML estimates of migration flows using the VenRePS Survey. The dependent variable is the number of migration observations, and the main explanatory variable is the standardized LQ correlation between municipalities. All specifications control for standardized driving distance, standardized absolute differences in ecological characteristics (rainfall, distance to coast, elevation, and temperature), and origin and destination municipality fixed effects. Column (2) additionally controls for standardized absolute difference in population density. Column (3) instead controls for the standardized Facebook Social Connectedness Index (SCI), while column (4) includes both population density and Facebook SCI. Standard errors are reported in parentheses and are clustered at the Venezuelan state–Colombian department pair level, where clusters are defined by unique combinations of Venezuelan state and Colombian department. Average marginal effects (AMEs) are reported for LQ correlation and Facebook SCI where applicable. The ratio is computed as the AME of LQ correlation divided by the AME of Facebook SCI. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

similarities. More specifically, we capture the correlation between LQ-measures of industrial presence for every origin-destination pair after removing one industry from the data. Having computed these measures for each of the 15 harmonized macro-industries, we re-run the PPML bilateral-flow gravity models. Figure 3 provides estimates of the β coefficient from our main specification (Column 3 of Table 1) after excluding each macro-industry from the data. Our results suggest that removing the agriculture sector does attenuate our results such that the coefficient is not statistically significant. Figure A.3 replicates this analysis on state-level data, iteratively excluding detailed industries that correspond to each of the 15 macro-industries considered at the municipality level. In this instance, excluding agriculture from the analysis does not attenuate the estimate but yields higher standard errors such that estimates for β are also not statistically significant. Interestingly, removing detailed industries associated with the mining macro-industry from this state-level analysis does attenuate the estimate. Overall, these results suggest that the agriculture sector plays an important role in explaining our main findings.

Figure 3: Migration Flows and LQ Similarity Excluding One Industry at a Time

(a) All controls

(b) Without population density control



Notes: The figure reports the estimated coefficients of the standardized LQ correlation from PPML regressions of migration flows from Venezuelan municipalities to Colombian municipalities (VenReps). Each coefficient is obtained after excluding the industry indicated on the vertical axis from the LQ industry vector before computing the correlation between each Venezuelan municipality and each Colombian municipality. Panel (a) reports the specification with all controls. Panel (b) reports the specification excluding the population density control. The specification otherwise follows the baseline municipality-to-municipality exercise reported in Table 1. Horizontal lines report 90% and 95% confidence intervals. The red dashed vertical line indicates zero.

5 Conclusion

Understanding how refugees sort across localities in host countries despite conditions limiting their agency is a crucial question for policymakers and academics at times of increased displacement, instability and political backlash. This paper provides evidence in favor of the "industrial matching" hypothesis in refugee settlement choices. We document this pattern in the context of the Venezuelan refugee shock, one of the largest processes of international displacement in recent memory. Moreover, we evaluate refugee settlement choices in Colombia, a developing country with limited resources and similar culture, ethnicity and shared history as Venezuela. Our results suggest that even in these conditions, displaced individuals sort towards labor markets where productive skills learned at home are transferable.

This study has several implications for policymakers concerned about refugees' economic integration. By showing that, even in contexts of forced displacement, migrants sort rationally towards areas that demand their productive skills, we provide governments with evidence that employment bans or restrictions may limit refugees' ability to begin integrating into labour markets and contribute to the economy (Fasani, Frattini and Minale, 2021; Marbach, Hainmueller and Hangartner, 2018). Moreover, being aware of how refugees make settlement choices allows for better planning by national and local governments to thus allocate resources efficiently as soon as refugee inflows start. If supported with information and unencumbered by regulation, refugees may organically reach the most productive opportunities to integrate and add to the economy.

Future research will need to build on these findings to evaluate whether they are externally valid and replicate in other environments, especially when refugees move towards more dissimilar countries with harsher attitudes against them. Moreover, as Colombia allowed Venezuelan refugees to easily regularize their legal situation in the country, it will be important to study whether obtaining legal status enables refugees to sort towards host-country localities with better employment opportunities for their specific set of skills.

References

Acharya, Avidit, Kirk Bansak and Jens Hainmueller. 2022. "Combining Outcome-Based and Preference-Based Matching: A Constrained Priority Mechanism." *Political Analysis* 30(1):89–112.

Adhikari, Prakash. 2013. "Conflict-induced displacement, understanding the causes of flight." *American Journal of Political Science* 57(1):82–89.

Ahani, Narges, Tommy Andersson, Alessandro Martinello, Alexander Teytelboym and Andrew C. Trapp. 2021. "Placement Optimization in Refugee Resettlement." *Operations Research* 69(5):1468–1486.

Aksoy, Cevat Giray, Panu Poutvaara and Felicitas Schikora. 2023. "First Time Around: Local Conditions and Multi-dimensional Integration of Refugees." *Journal of Urban Economics* 137:103588.

Åslund, Olof and Dan-Olof Rooth. 2007. "Do When and Where Matter? Initial Labour Market Conditions and Immigrant Earnings." *The Economic Journal* 117(518):422–448.

Bahar, Dany, Ana María Ibáñez and Sandra V. Rozo. 2021. "Give me your tired and your poor: Impact of a large-scale amnesty program for undocumented refugees." *Journal of Development Economics* 151:102652.
URL: *https://www.sciencedirect.com/science/article/pii/S0304387821000316*

Bansak, Kirk, Jeremy Ferwerda, Jens Hainmueller, Andrea Dillon, Dominik Hangartner, Duncan Lawrence and Jeremy Weinstein. 2018. "Improving Refugee Integration through Data-Driven Algorithmic Assignment." *Science* 359(6373):325–329.

Bazzi, Samuel, Arya Gaduh, Alexander D Rothenberg and Maisy Wong. 2016. "Skill transferability, migration, and development: Evidence from population resettlement in Indonesia." *American Economic Review* 106(9):2658–2698.

Blouin, Cécile, Isabel Berganza and Luisa Feline Freier. 2020. "The spirit of Cartagena? Applying the extended refugee definition to Venezuelans in Latin America.".

Borjas, George J. 2001. "Does Immigration Grease the Wheels of the Labor Market?" *Brookings Papers on Economic Activity* 2001(1):69–133.

Conte, Alessandra and Silvia Migali. 2019. "The role of conflict and organized violence in international forced migration." *Demographic Research* 41:393–424.

Corrales, Javier. 2023. *Autocracy rising: How Venezuela transitioned to authoritarianism.* Bloomsbury Publishing PLC.

Correia, Sergio, Paulo Guimarães and Tom Zylkin. 2020. "Fast Poisson estimation with high-dimensional fixed effects." *The Stata journal* 20(1):95–115.

Damm, Anna Piil. 2009. "Ethnic Enclaves and Immigrant Labor Market Outcomes: Quasi-Experimental Evidence." *Journal of Labor Economics* 27(2):281–314.

DANE. 2025. "Encuesta Pulso de la Migración (EPM), Rondas 1–8, 2021–2025." Dataset. **URL:** *https://www.dane.gov.co/index.php/estadisticas-por-tema/demografia-y-poblacion/encuesta-pulso-de-la-migracion-epm*

de Haas, Hein, Mathias Czaika, Marie-Laurence Flahaux, Edo Mahendra, Katharina Natter, Simona Vezzoli and María Villares-Varela. 2019. "International Migration: Trends, Determinants, and Policy Effects." *Population and development review* 45(4):885–922.

Departamento Nacional de Planeación. 2025. "Evaluación institucional del Documento CONPES 4100 de 2022 y de resultados de las iniciativas adelantadas para la atención e integración de la población migrante venezolana." Archivo Nacional de Datos (ANDA). **URL:** *https://anda.dnp.gov.co/index.php/catalog/176/overview*

Edin, Per-Anders, Peter Fredriksson and Olof Åslund. 2003. "Ethnic Enclaves and the

Economic Success of Immigrants—Evidence from a Natural Experiment." *The Quarterly Journal of Economics* 118(1):329–357.

Fafchamps, Marcel and Forhad Shilpi. 2013. "Determinants of the choice of migration destination." *Oxford Bulletin of Economics and Statistics* 75(3):388–409.

Fasani, Francesco, Tommaso Frattini and Luigi Minale. 2021. "Lift the Ban? Initial Employment Restrictions and Refugee Labour Market Outcomes." *Journal of the European Economic Association* 19(5):2803–2854.

Freier, Luisa Feline, Isabel Berganza and Cécile Blouin. 2022. "The Cartagena Refugee Definition and Venezuelan Displacement in Latin America." *International migration* 60(1):18–36.

Freier, Luisa Feline and Leda M. Pérez. 2021. "Nationality-Based Criminalisation of South-South Migration: the Experience of Venezuelan Forced Migrants in Peru." *European journal on criminal policy and research* 27(1):113–133.

Freier, Luisa Feline and Nicolas Parent. 2019. "The Regional Response to the Venezuelan Exodus." *Current History* 118(805):pp. 56–61.
URL: *https://www.jstor.org/stable/48614415*

Gottwald, Martin. 2004. "Protecting Colombian refugees in the Andean region: the fight against invisibility." *International Journal of Refugee Law* 16(4):517–546.

Hammoud-Gallego, Omar. 2024. "The short-term effects of visa restrictions on migrants' legal status and well-being: A difference-in-differences approach on Venezuelan displacement." *World Development* 182:106709.
URL: *https://linkinghub.elsevier.com/retrieve/pii/S0305750X24001797*

Hammoud-Gallego, Omar and Luisa Feline Freier. 2023. "Symbolic Refugee Protection:

Explaining Latin America's Liberal Refugee Laws." *The American political science review* 117(2):454–473.

Ibáñez, Ana María, Andrés Moya and Sandra V. Rozo. 2021. "Venezuelan Refugee Panel Study (VenRePS)." Dataset. Universidad de los Andes, Inter-American Development Bank, and World Bank.
URL: *https://www.venreps.com/*

Ibáñez, Ana María, Andrés Moya, María Adelaida Ortega, Sandra V Rozo and Maria José Urbina. 2025. "Life Out of the Shadows: the Impacts of Regularization Programs on the Lives of Forced Migrants." *Journal of the European Economic Association* 23(3):941–982.
URL: *https://doi.org/10.1093/jeea/jvae044*

Ibáñez, Ana María, Sandra V. Rozo, Dany Bahar and María José Urbina. 2026. "Protecting the vulnerable: How migrant regularization reduces crime and empowers women." *Journal of Development Economics* 179:103667.
URL: *https://www.sciencedirect.com/science/article/pii/S0304387825002184*

Johnston, Drew, Theresa Kuchler, Manas Kulkarni and Johannes Stroebel. 2026. "The Social Connectedness Index: A large-scale dataset of social ties across geographic locations." *Data in Brief* p. 112905.

Kang, Yoo-Duk. 2021. "Refugee crisis in Europe: Determinants of asylum seeking in European countries from 2008–2014." *Journal of European Integration* 43(1):33–48.

Kennan, John and James R. Walker. 2011. "The Effect of Expected Income on Individual Migration Decisions." *Econometrica* 79(1):211–251.

Knight, Brian and Ana Tribin. 2023. "Immigration and violent crime: Evidence from the Colombia-Venezuela border." *Journal of Development Economics* 162:103039.

Lesser, Jeff. 2013. *Immigration, ethnicity, and national identity in Brazil, 1808 to the present.* New approaches to the Americas Cambridge: Cambridge University Press.

Lewis, Benjamin and Devika Jain. 2016. "Harvard CGA Geotweet Archive v2.0.".
URL: *https://doi.org/10.7910/DVN/3NCMB6*

Marbach, Moritz, Jens Hainmueller and Dominik Hangartner. 2018. "The Long-Term Impact of Employment Bans on the Economic Integration of Refugees." *Science Advances* 4(9):eaap9519.

Marques, Renan. 2025. "Forced migration and violent crime: Evidence from the Venezuelan exodus to Brazil." *The Journal of Development Studies* 61(2):210–232.

Matsui, Nozomi and James Raymer. 2020. "The push and pull factors contributing towards asylum migration from developing countries to developed countries since 2000." *International Migration* 58(6):210–231.

Morales-Arilla, Jose, Diego Martin and Alvaro Morales. 2025. "Escaping from hardship, searching for comfort: Climate matching in refugees' destination choices." *Journal of Development Economics* 176:103500.

Munshi, Kaivan. 2003. "Networks in the Modern Economy: Mexican Migrants in the U.S. Labor Market." *The Quarterly Journal of Economics* 118(2):549–599.

Obolensky, Marguerite, Marco Tabellini and Charles Taylor. 2024. "Homeward Bound: How Migrants Seek Out Familiar Climates.".
URL: *http://www.nber.org/papers/w32035*

on Foreign Relations, Council. 2024. "Venezuela: The Rise and Fall of a Petrostate.".
URL: *https://www.cfr.org/backgrounders/venezuela-crisis*

Patt, Alexander, Jens Ruhose, Simon Wiederhold and Miguel Flores. 2021. "International

Emigrant Selection on Occupational Skills." *Journal of the European Economic Association* 19(2):1249–1298.

R4V. 2026. "Refugees and Migrants from Venezuela.".
URL: *https://www.r4v.info/en/refugeeandmigrants*

Regional Inter-Agency Coordination Platform for Refugees and Migrants from Venezuela. 2024. "Refugees and Migrants from Venezuela." R4V.
URL: *https://www.r4v.info/en/refugeeandmigrants*

Spörlein, Christoph. 2015. "Destination choices of recent Pan–American migrants: opportunities, costs, and migrant selectivity." *International Migration Review* 49(2):523–552.

Stuart, Bryan A and Evan J Taylor. 2021. "Migration networks and location decisions: Evidence from US mass migration." *American Economic Journal: Applied Economics* 13(3):134–175.

Tubadji, Annie and Peter Nijkamp. 2015. "Cultural gravity effects among migrants: A comparative analysis of the EU15." *Economic Geography* 91(3):343–380.

UNHCR. 2026. Global Report 2025. Technical report United Nations High Commissioner for Refugees.
URL: *https://www.unhcr.org/media/global-report-2025*

Unidad Administrativa Especial Migración Colombia. 2021. "Política Institucional de Derechos Humanos.".

Van Hear, Nicholas, Oliver Bakewell and Katy Long. 2018. "Push-pull plus: reconsidering the drivers of migration." *Journal of ethnic and migration studies* 44(6):927–944.

Wang, Zhiling, Thomas De Graaff and Peter Nijkamp. 2016. "Cultural diversity and cultural distance as choice determinants of migration destination." *Spatial Economic Analysis* 11(2):176–200.

Online Appendix

Supplementary Materials for the research article "Industrial Matching in Refugees' Settlement Choices"

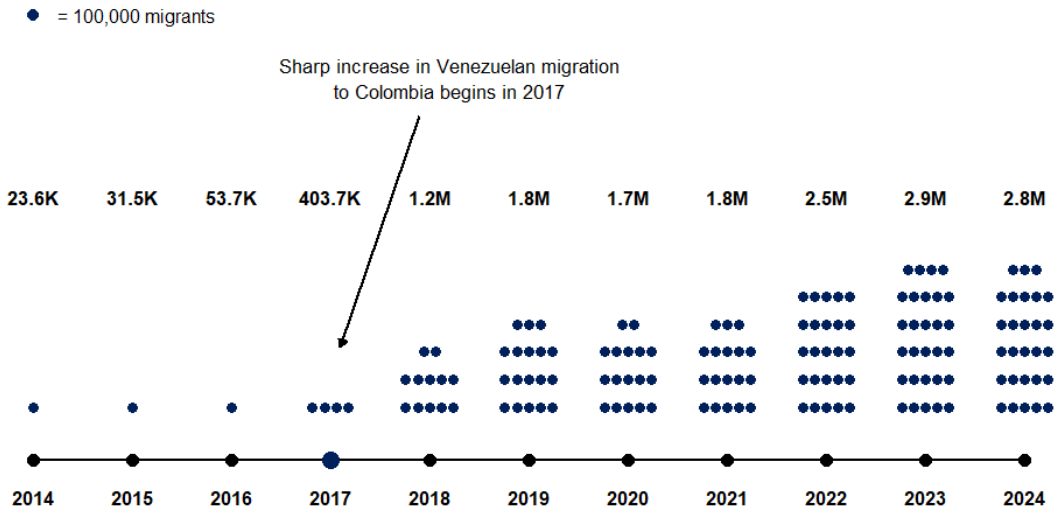
Additional Analyses

August 10, 2026

Table of Contents

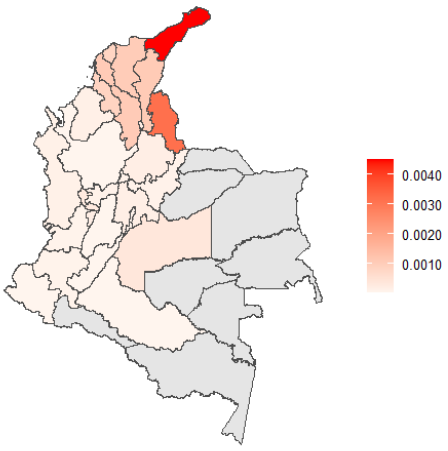
A Additional Figures

Figure A.1: Venezuelan Migration to Colombia, 2014–2024

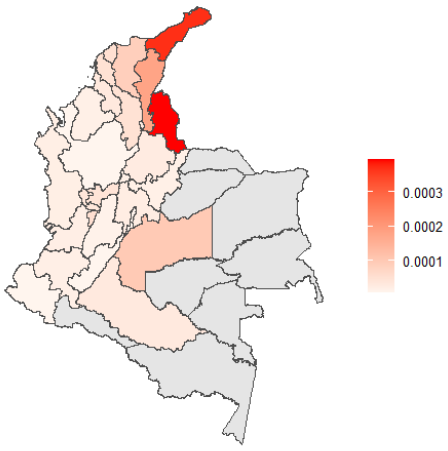


Notes: Each blue dot represents approximately 100,000 Venezuelan migrants in Colombia. For 2014–2016, the reported population is below 100,000; therefore, a single dot is shown for each of these years. R4V figures correspond to December 2021, December 2022, November 2023, and November 2024.
Source: Authors' elaboration based on (Departamento Nacional de Planeación, 2025) for 2014; (Unidad Administrativa Especial Migración Colombia, 2021) for 2015–2020; and the (Regional Inter-Agency Coordination Platform for Refugees and Migrants from Venezuela, 2024) for 2021–2024.

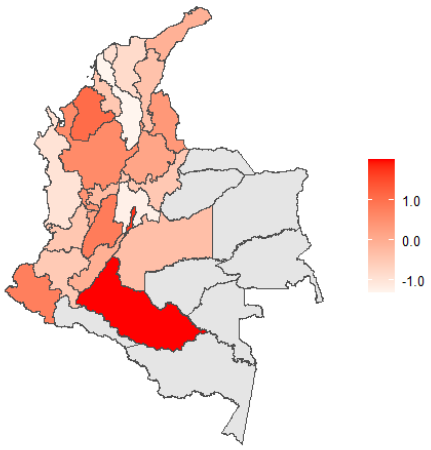
Figure A.2: VenRep Flows and Employment Similarities at the State–Department Level



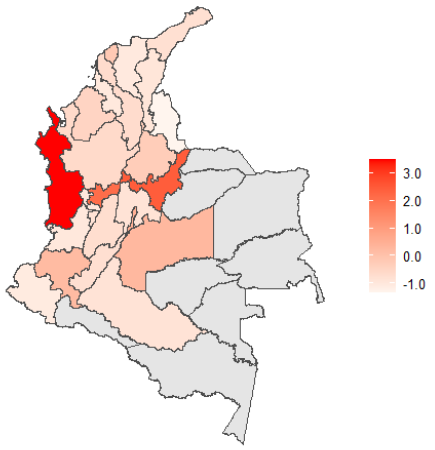
(a) Caracas migrants presence (VenReps)



(b) Bolívar migrants presence (VenReps)



(c) Caracas employment similarity

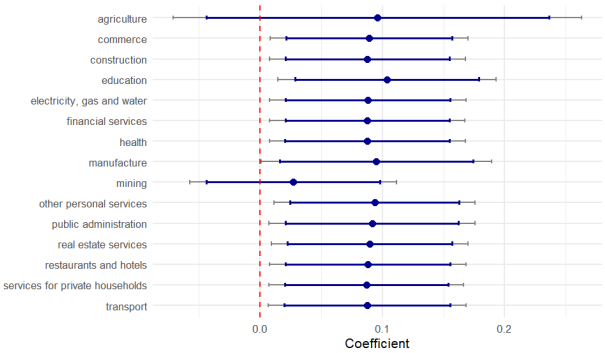


(d) Bolívar employment similarity

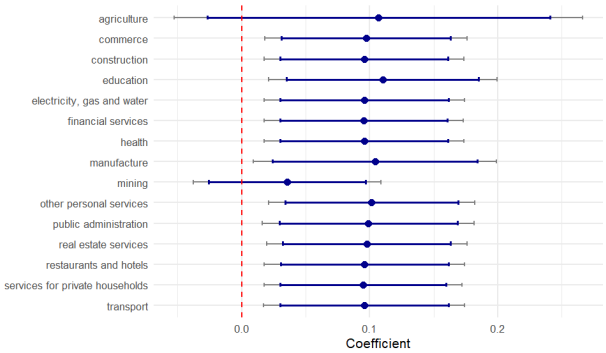
Notes: VenRep flows correspond to the inverse hyperbolic sine transformation (`asinh`) of the number of VenRep flows per 1,000 inhabitants in each Colombian department. Employment similarities correspond to the standardized revealed comparative advantage LQ similarity measure between Venezuelan states and Colombian departments.

Figure A.3: State-Level Migration Flows and LQ Similarity Excluding Sectors at a Time

(a) All controls



(b) Without population density control



Notes: The figure reports the estimated coefficients of the standardized LQ correlation from PPML regressions of migration flows from Venezuelan states to Colombian departments (VenReps). Each coefficient is obtained after excluding the industries within the macro-industry indicated on the vertical axis from the LQ industry vector before computing the correlation between each Venezuelan state and each Colombian department. Panel (a) reports the specification with all controls. Panel (b) reports the specification excluding the population density control. The specification otherwise follows the baseline exercise reported in Table 1 but at a state-to-department. Horizontal lines report 90% and 95% confidence intervals. The red dashed vertical line indicates zero.

B Additional Tables

Table A.1: Summary Statistics at the Municipality Level

Variable	N	Mean	SD	Min	1st Qu.	Median	3rd Qu.	Max
Outcomes								
VenReps survey migrant observations	166332	2.66	60.90	0.00	0.00	0.00	0.00	11539.00
Migration Pulse	177215	8.71	271.55	0.00	0.00	0.00	0.00	74208.49
Number of Twitter/X users	177215	0.13	3.89	0.00	0.00	0.00	0.00	1108.00
Independent variables								
Standardized *LQ* correlation	177215	0.00	1.00	-2.56	-0.69	-0.19	0.79	1.97
Standardized industry-share correlation	177215	0.00	1.00	-2.22	-0.94	0.09	0.94	1.51
Controls								
Standardized driving distance	177215	0.00	1.00	-6.11	-0.58	0.10	0.70	2.31
Standardized rainfall	177215	0.00	1.00	-6.77	-0.34	0.26	0.67	1.03
Standardized distance to coast	177215	0.00	1.00	-5.93	-0.51	0.22	0.76	1.20
Standardized elevation	176880	0.00	1.00	-2.77	-0.73	0.18	0.88	1.31
Standardized temperature	177215	0.00	1.00	-3.42	-0.66	0.19	0.85	1.38
Standardized population density	177215	0.00	1.00	-9.77	0.09	0.34	0.40	0.43

Notes: This table reports summary statistics at the level of municipality pairs between Venezuelan municipalities and Colombian municipalities. Control variables were standardized after aggregation because the original ecological similarities dataset only contained standardized pairwise differences rather than municipality-level raw values.

Table A.2: Summary Statistics at the State–Department Level

Variable	N	Mean	SD	Min	1st Qu.	Median	3rd Qu.	Max
Outcomes								
VenReps survey migrant observations	576	706.58	3002.11	0.00	22.75	102.00	379.75	56261.00
Migration Pulse	576	2661.93	7498.14	0.00	0.00	411.41	1754.90	74208.49
Number of Twitter/X users	576	38.12	106.91	0.00	2.00	6.00	25.00	1186.00
Independent variables								
Standardized *LQ* correlation	576	0.00	1.00	-1.73	-0.68	-0.20	0.49	4.21
Standardized industry-share correlation	576	0.00	1.00	-2.80	-0.72	0.27	0.82	1.50
Controls								
Standardized driving distance	576	0.00	1.00	-3.55	-0.59	0.11	0.72	2.25
Standardized rainfall	576	0.00	1.00	-4.70	-0.29	0.29	0.62	1.10
Standardized distance to coast	576	0.00	1.00	-4.41	-0.43	0.30	0.72	1.10
Standardized elevation	576	0.00	1.00	-3.41	-0.60	0.08	0.83	1.58
Standardized temperature	576	0.00	1.00	-3.24	-0.66	0.12	0.82	1.60
Standardized population density	576	0.00	1.00	-4.20	0.17	0.34	0.38	0.43

Notes: This table reports summary statistics at the level of pairs between Venezuelan states and Colombian departments. Control variables were standardized after aggregation because the original ecological similarities dataset only contained standardized pairwise differences rather than state- or department-level raw values.

Table A.3: Venezuelan to Colombian Municipalities: Migration Flows and S Similarity

	(1)	(2)	(3)
Industrial Similarity - S (sd)	0.186**	0.294***	0.256***
	(0.082)	(0.047)	(0.036)
AME Ind. Similarity (sd)	0.495**	0.783***	0.682***
	(0.196)	(0.115)	(0.088)
AME Driving distance (sd)	–	7.267***	4.000***
	–	(1.722)	(1.327)
AME Ratio: Ind. Sim. / Distance	–	0.108	0.171
Observations	166,332	166,332	165,998
Origin Muni. FE	Yes	Yes	Yes
Dest. Muni. FE	Yes	Yes	Yes
Driving distance	No	Yes	Yes
Ecological controls	No	No	Yes

Notes: The table reports PPML estimates of migration flows from Venezuelan municipalities to Colombian municipalities using the VenReps Survey. The dependent variable is the number of migration observations. The main explanatory variable is the standardized correlation of employment shares S across municipalities. Columns progressively add controls: (2) standardized driving distance and (3) standardized absolute differences in ecological characteristics (rainfall, distance to coast, elevation, and temperature). Standard errors are reported in parentheses and are clustered at the Venezuelan state–Colombian department pair level, where clusters are defined by unique combinations of Venezuelan state and Colombian department. Average marginal effects (AME) are reported for employment structure similarity and driving distance. The ratio is computed as the AME of employment shares S similarity divided by the AME of driving distance. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.4: State-level data: Migration Flows and LQ Similarity

	(1)	(2)	(3)
Industrial Similarity - LQ (sd)	0.087**	0.112**	0.096**
	(0.044)	(0.053)	(0.040)
AME Ind. Similarity (sd)	61.653**	79.007**	67.620**
	(30.858)	(37.665)	(27.997)
Observations	576	576	576
Origin Region FE	Yes	Yes	Yes
Dest. Region FE	Yes	Yes	Yes
Driving distance	No	Yes	Yes
Ecological controls	No	No	Yes

Notes: The table reports PPML estimates of migration flows from Venezuelan states to Colombian departments using the VenReps Survey. The dependent variable is the number of migration observations. The main explanatory variable is the standardized LQ correlation between regions. Columns progressively add controls: (2) standardized driving distance, (3) standardized absolute differences in ecological characteristics (rainfall, distance to coast, elevation, and temperature). Standard errors are reported in parentheses and are two-way clustered at the Colombian department and Venezuelan state levels. Average marginal effects (AME) are reported for LQ similarity and driving distance. The ratio is computed as the AME of LQ similarity divided by the AME of driving distance. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.5: Municipality-level LQ Similarity Across Migration Data Sources

	VenRePS	Pulso Migratorio	Twitter/X
Industrial Similarity - LQ (sd)	0.090**	0.007	0.052**
	(0.040)	(0.064)	(0.021)
AME Ind. Similarity (sd)	0.240**	0.207	0.010**
	(0.100)	(1.979)	(0.004)
AME Driving distance (sd)	3.802***	94.267***	0.218***
	(1.238)	(13.818)	(0.060)
AME Ratio: Ind. Sim. / Distance	0.063	0.002	0.046
Observations	165,998	50,320	117,160
Origin Muni. FE	Yes	Yes	Yes
Dest. Muni. FE	Yes	Yes	Yes
Driving distance	Yes	Yes	Yes
Ecological controls	Yes	Yes	Yes

Notes: The table compares municipality-to-municipality PPML estimates of migration flows as dependent variables across three migration datasets: VenRePS, Pulso Migratorio, and Twitter/X. The main explanatory variable is the standardized LQ correlation between Venezuelan and Colombian municipalities. All specifications include standardized driving distance, standardized absolute differences in ecological characteristics (rainfall, distance to coast, elevation, and temperature), as well as origin and destination municipality fixed effects. Standard errors are reported in parentheses and are clustered at the Venezuelan state–Colombian department pair level, where clusters are defined by unique combinations of Venezuelan state and Colombian department. Average marginal effects (AME) are reported for LQ correlation and driving distance. The ratio is computed as the AME of LQ correlation divided by the AME of driving distance. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.6: State-level LQ Similarity Across Migration Data Sources

	VenRePS	Pulso Migratorio	Twitter/X
Industrial Similarity - LQ (sd)	0.096**	0.162***	0.052*
	(0.040)	(0.046)	(0.030)
AME Ind. Similarity (sd)	67.620**	430.105***	1.995*
	(27.997)	(127.217)	(1.162)
Observations	576	576	576
Origin Region FE	Yes	Yes	Yes
Dest. Region FE	Yes	Yes	Yes
Driving distance	Yes	Yes	Yes
Ecological controls	Yes	Yes	Yes

Notes: The table compares Venezuelan States-to-Colombian Departments PPML estimates of migration flows as dependent variables across three migration datasets: VenRePS, Pulso Migratorio, and Twitter/X. The main explanatory variable is the standardized LQ correlation across Venezuelan states and Colombian departments. All specifications include standardized driving distance, standardized absolute differences in ecological characteristics (rainfall, distance to coast, elevation, and temperature), as well as origin and destination municipality fixed effects. Standard errors are reported in parentheses and are clustered at the Venezuelan state–Colombian department pair level, where clusters are defined by unique combinations of Venezuelan state and Colombian department. Average marginal effects (AME) are reported for LQ correlation and driving distance. The ratio is computed as the AME of LQ correlation divided by the AME of driving distance. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.