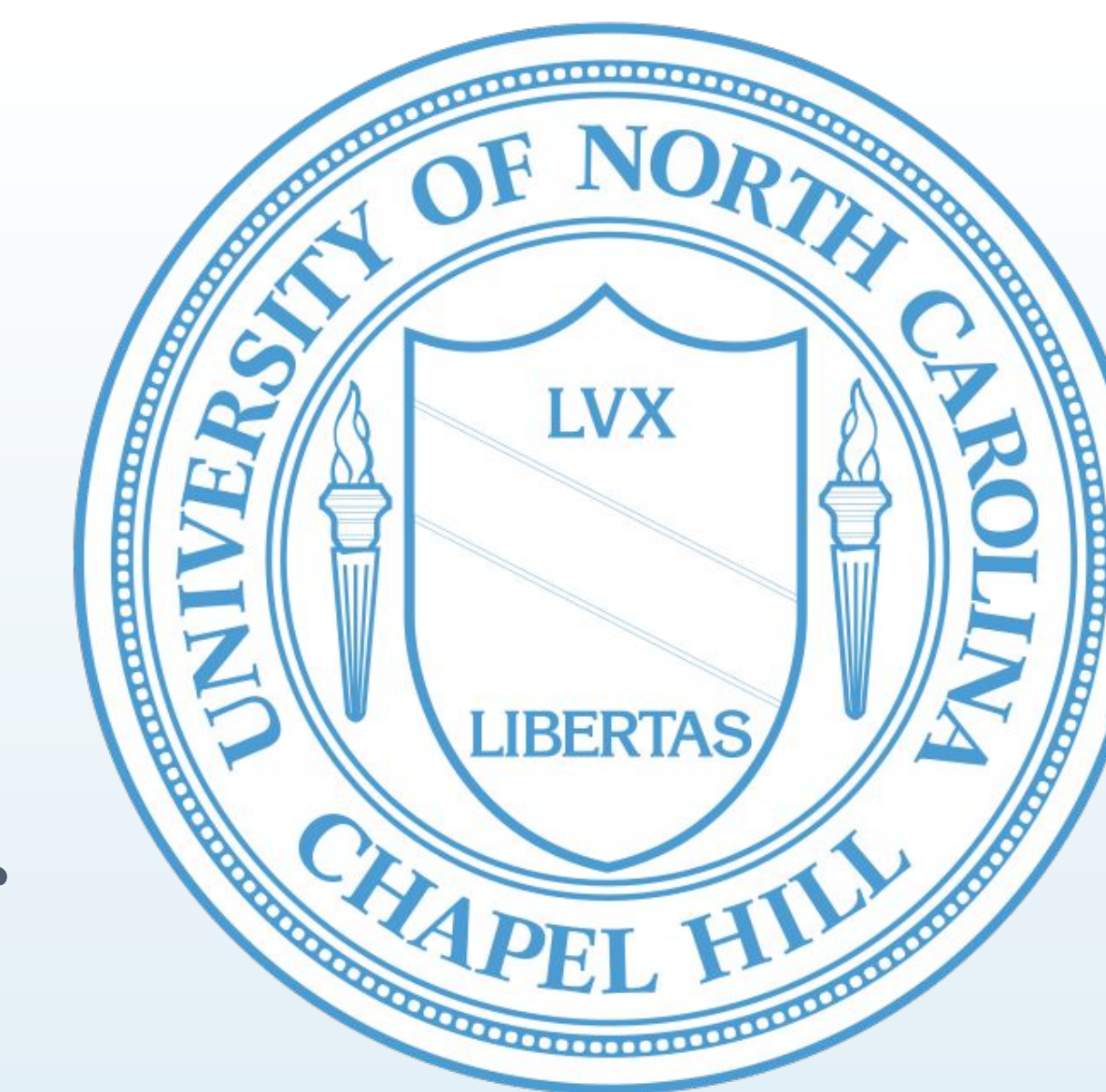


Modeling Migration: A Data-Driven Analysis of Displacement in Venezuela and Choice in Colombia

Investigation in Economics by Jose Guevara Hernandez under the direction of Dr. Valentin Verdier
MURAP 2023 Summer Fellowship



ABSTRACT

The location where a migrant settles is a consequential factor in their economic success. Each migrant is different, and each location is different, so it follows that some migrant/location pairs result in better outcomes than others. Market design principles can be used to develop algorithms that allocate migrants across available locations via outcome-based matching, preference-based matching, or a combination of both. Since there is no systemic data on migrant preferences, these have to be inferred. This paper introduces a methodology to derive the preferences of Venezuelan migrants over departments in Colombia. We create a model to evaluate how well preferences can be ascertained from migrants' realized choices in location. Using historical data on Venezuelan migrants and the departments in Colombia where they settle, we analyze the effect of family-specific variables (e.g., number of children) and department-specific variables (e.g., educational achievement) to see what role these play in determining preferences. Results suggest that historical data is suboptimal for inferring preferences, indicating that migrant choices might be determined by something extraneous to the data, such as family ties or income constraints.

BACKGROUND

Venezuela is home to one of the largest displacement emergencies in the globe (CDP). As a result of economic collapse in the country, millions have been without access to food or even medicine (Amnesty International). As of 2023, over 7.24 million Venezuelans have fled the country of which Colombia hosts over 2.5 million (R4V).

Currently, the bulk of displaced Venezuelans are concentrated in a few border municipalities. This uneven distribution places strain on the infrastructure of the border municipalities, which do not have the resources to handle such large population influxes (Bahar). Hospitals and schools are at capacity in these border regions, while larger metropolitan areas have much lower migrant concentrations (Moloney).

To address this issue, many have proposed redistribution schemes in which migrants are "matched" to a certain location. **The question is, who goes where, and why?**

METHODOLOGY

Individuals have preferences regarding where they want to live. A migrant family's choice in location can then be seen as a utility maximization problem, denoted as:

$$\text{realized location} = j^* = \arg \max_{j=1, \dots, J} U_{ij}$$

The utility obtained by moving to department j is determined by x_i , a vector of family-specific characteristics interacted with z_j , a vector of location-specific characteristics, and χ_j , an intercept, capturing the average utility for each location. We denote it:

$$U_{ij} = \chi_j + \beta(x_i \cdot z_j) + \varepsilon_{ij}$$

The probability that family i will choose department j^* over all other departments is denoted:

$$P(j^* | x_i, z_j : j = 1, \dots, J) = \frac{\exp(\chi_j + \beta(x_i \cdot z_j))}{\sum_{j=1, \dots, J} \exp(\chi_j + \beta(x_i \cdot z_j))}$$

We use Maximum Likelihood Estimation to find the parameter values that maximize the likelihood that our function will match the observed outcome.

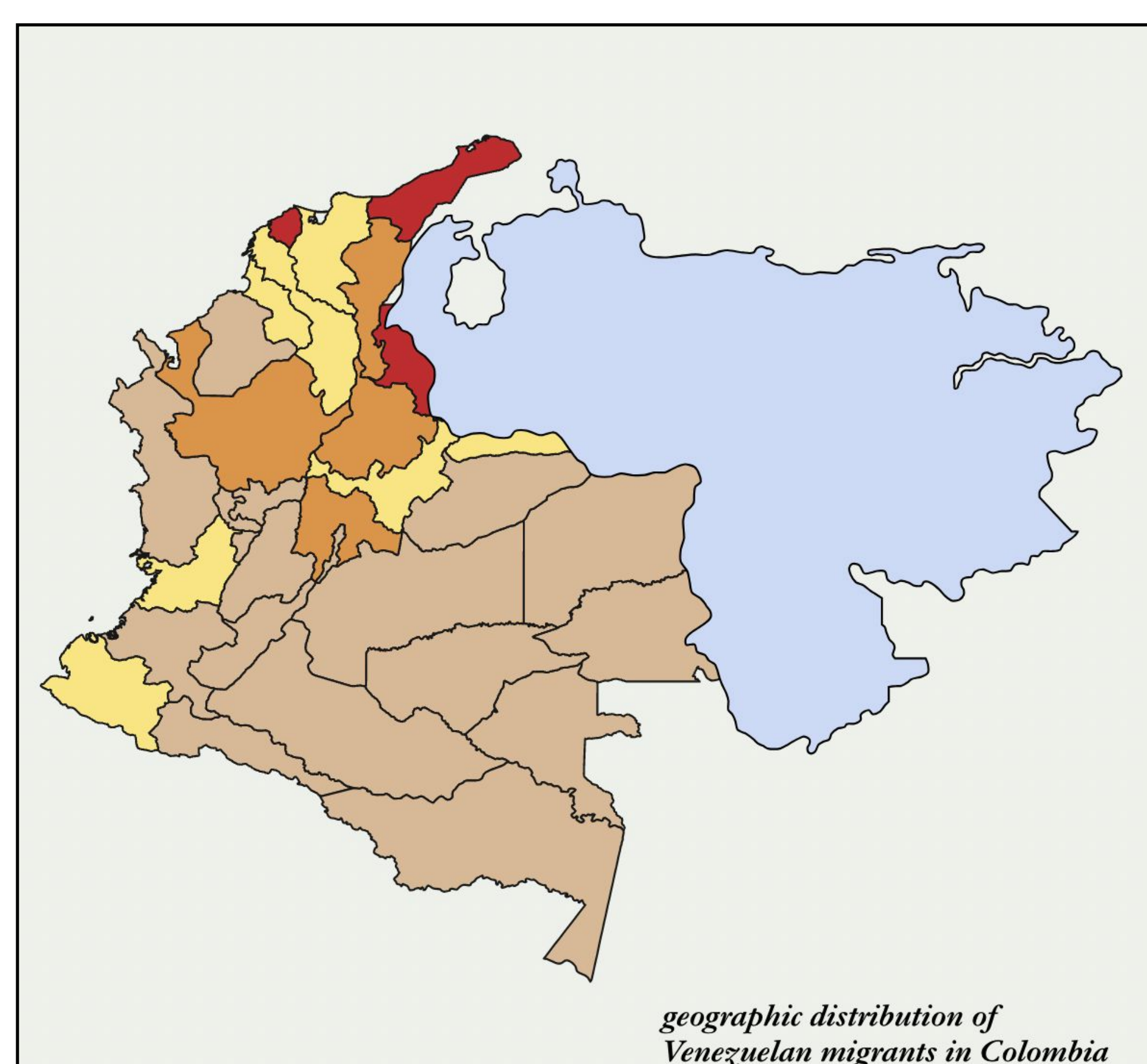
RESULTS

Table 3: Logit Model for Realized Choice

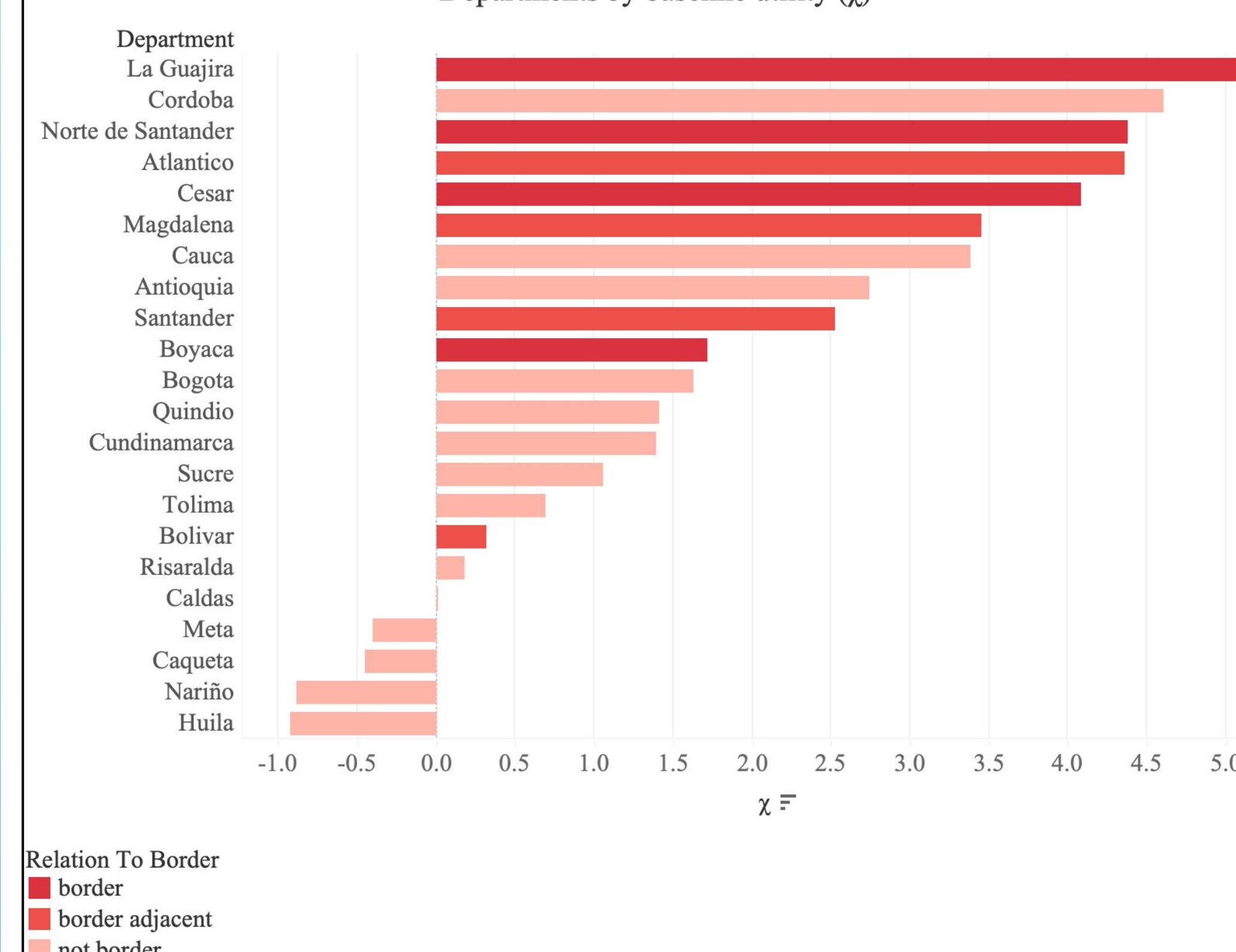
	Model	Border Model [†]
Department		
schoolkids#underachievement	0.017*** (0.00425)	0.013 (0.0122)
elders#healthbarrier	0.017* (0.00665)	0.045** (0.0142)
edulevel#supplyratio	-0.52*** (0.0544)	-0.69*** (0.191)
Bogota		
_cons	1.632* (0.636)	
Antioquia		
_cons	2.747*** (0.215)	
Atlantico		
_cons	4.365*** (0.421)	
Norte de Santander		
_cons	4.381*** (0.420)	2.826*** (0.419)
Valle del Cauca		
_cons	1.164* (0.589)	
<i>19 rows omitted</i>		
<i>N</i>	5,077	1,629
<small>standard errors in parentheses * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ [†] model that omits non-border departments</small>		

The coefficients on the interaction between number of kids low achievement, as well as number of elderly family members and health barriers, are positive. The coefficient for the interaction between highest level of education achieved and job availability is negative.

MAP



Departments by baseline utility (χ)



The model estimates χ , a measure of baseline utility, for each location. This bar chart shows each department's χ value, color-coded to reflect proximity to the border with Venezuela. As we can see, border departments are associated with higher levels of utility.

FINDINGS

The relationship observed by the data seems counterintuitive. Why would a family with more children prefer a location with worse quality education? Or why would elderly people move to locations with worse healthcare? We consider two possibilities.

Missing Variables

It is possible that the model omits critical variables that influence choice. Factors like family connections may play a large role, but they are challenging to quantify and include in the analysis. Similarly, income constraints could influence the ability of migrants to move to locations they prefer. Thus **the data may not truly reflect preferences but instead capability.**

General Limitations of Choice Models

Choice models require that an agent's preferences are complete and transitive and **assume that they can systematically rank all options.** This may not be realistic, as migrant families likely do not have complete information on all departments. It is also possible that families are not considering optimal conditions but are simply settling in the first location they can move to.

AVENUES FOR FUTURE RESEARCH

1. Analysis with expanded data, including information on points of origin *within* Venezuela, income, previous occupation/labor skills.
2. Surveys of migrant preferences and motivations or Department "pull factors".
3. Modeling preferences of migrants informed on the profiles of each Department.

ACKNOWLEDGEMENTS

I would like to thank the following for their support and guidance throughout this Investigation

- **Dr. Valentin Verdier**, Mentor
- **Matt Jansen**, Data Cleaning and Analysis
- **Alejandro Cotté Alsina**, Co-mentee,
- **Justin Hadad**,
- **Dr. Gina Galindo**,
- **The MURAP Scholars Program and staff**

Table 3: Logit Model for Realized Choice

	Model	Border Model [†]
Department		
schoolkids#underachievement	0.017*** (0.00425)	0.013 (0.0122)
elders#healthbarrier	0.017* (0.00665)	0.045** (0.0142)
edulevel#supplyratio	-0.52*** (0.0544)	-0.69*** (0.191)
Bogota		
_cons	1.632* (0.636)	
Antioquia		
_cons	2.747*** (0.215)	
Atlantico		
_cons	4.365*** (0.421)	
Norte de Santander		
_cons	4.381*** (0.420)	2.826*** (0.419)
Valle del Cauca		
_cons	1.164* (0.589)	
<i>19 rows omitted</i>		
<i>N</i>	5,077	1,629
standard errors in parentheses		
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$		
[†] model that omits non-border departments		