Self-selection of Venezuelan Migrants: Who Migrates Where? *

Federico Maggio†

August 20, 2022

Abstract

I provide the first evidence on self-selection of Venezuelan migrants who arrived in Colombia, Peru, Chile, and the United States over the 2015-2018 period. The analysis relies on individual-level data for the Venezuelan population residing in the origin country and in the four main destination countries. I find that both male and female migrants are positively selected compared to stayers in terms of education. Moreover, South-North migrants turn out to be positively sorted with respect to South-South migrants. However, I provide evidence that controlling for self-selection of unobservables, only migrants residing in Chile and the United States are positively selected in terms of expected income, whereas South-South migrants appear negatively selected. I provide an interpretation of these patterns based on differences in the migration costs according to individual characteristics and destination countries' immigration policies.

JEL codes: F22, O15, J15, J16, J24
Keywords: Venezuela, Migration, Selection, Individual-level data, Education

*First of all, I wish to thank ... for the constructive comments and the useful suggestions provided.
†Free University of Bozen. Email: federico.maggio@economics.unibz.it.

1 Introduction

Since 2010, Venezuela has been undergoing the worst socioeconomic and political crisis ever experienced by any Latin American country (Bull and Rosales, 2020). The financial and economic crisis has led to a harsh humanitarian crisis, prompting an unprecedented wave of international emigration, with nearly 5.2 million Venezuelans leaving over the 2013-2019 period, from a country with a total population of 28.9 million. Although a substantial number of works have addressed the impact of Venezuelan migrants on the society and economy of neighboring countries (Olivieri et al., 2022; Rozo and Vargas, 2021), the characteristics of those who are leaving the country and the determinants driving them to choose a specific country have not yet been investigated.

Studying migrants' decision-making process, their self-selection according to sociodemographic characteristics, and intended destination countries is important for several reasons. First, understanding migrants' motivations helps to distinguish migrants forced to leave their countries from migrants seeking better opportunities. Second, the skill composition of migrants has critical implications for the reconstruction of the country of origin. The more skilled migrants are, the more difficult will be to rebuild the country at the end of the political and economic crisis. Third, knowing the level of intrinsic income determinants of migrants is useful for planning integration policies, thus contributing to the social stability of receiving countries. Such aspect is particularly relevant in episodes of migration toward developing countries (South-South migration), which even before the migration crisis were not experiencing stable political and economic conditions.

In the present study, migrants' self-selection is assessed in terms of observable individual characteristics, as well as predicted earnings. I also estimate the role of expected earnings and country-specific migration costs in shaping the individual migration decision, in an episode in which South-South migration outweighs South-North migration.

The current analysis relies on individual-level data for the Venezuelan population residing in the origin country, Colombia, Peru, Chile, and the US. To shed light on this issue, I follow three distinct steps. First, I estimate a linear probability model for each destination country related to how migrants differ from stayers in terms of observable characteristics. Second, to evaluate self-selection in terms of predicted income, cumulative distribution functions CDFs) of migrant's counterfactual earnings in the origin country are compared with stayers' CDFs. Since migrants are self-selected in their unobserved characteristics,

considering exclusively observable characteristics to predict their counterfactual earnings, had they chosen not to migrate, would lead to biased results. To overcome this issue, I estimate corrected counterfactual earnings by applying the Dahl (2002) self-selection correction method. Finally, to assess the role of expected gains and migration costs in shaping the individual migration decision, I estimate a mixed multinomial logit model using the counterfactual expected earnings for all alternative locations.

The first set of results shows that educated people are significantly more likely to migrate, indicating that migrants are positively selected in education compared to stayers. However, for both males and females, the effect of having a college degree on the probability of migrating is different depending on the country of destination. Indeed, self-selection is stronger in South-North migration than in South-South migration.

Second, I find that both genders of migrants in the United States and only females in Chile are positively selected in terms of expected income. Migrants in Colombia and Peru, on the other hand, seem to be negatively selected over the stayers.

The latest set of results shows that even controlling for expected earnings at destination, individual characteristics have a significant effect on the likelihood of migrating to a specific country. Such a result suggests that migration costs are highly dependent on the destination country and individual characteristics. Finally, my estimates show that migration costs related to the destination country in a South-North migration, are about three times higher than in a South-South migration.

Although there is extensive literature on migrant self-selection, the focus has been on South-North migration. Borjas (1990) laid the groundwork for the self-selection theory, showing that migrants from a country with high return to education inequality to a more equal country in terms of education premium are expected to come from the upper end of the skill distribution. Chiquiar and Hanson (2005), merging information from the US census on Mexican migrants' characteristics with information from the Mexican census on the stayers' characteristics, pioneered comparing the counterfactual wages of migrants in the origin country (as if they had not migrated) to the wages of stayers. They show that Mexican-born men were intermediately selected and Mexican-born women were positively selected. Grogger and Hanson (2011) confirmed the hypothesis that countries with high returns to skills attract highly skilled migrants. Using aggregate data on emigrant stocks by schooling level and source country in OECD destinations, they estimated an income maximization model and showed that migrants were positively selected and that more educated migrants

were more likely to settle in destination countries with high skill returns. However, none of these studies addressed how the distribution of unobserved characteristics may affect the probability of migration and wages at destination. Moraga (2011) and Kaestner and Malamud (2014), analyzing the Mexican migration to the US and using data reporting pre-migration wages, found negative selection of Mexican migrants. They also provided evidence that part of the negative selection could be attributed to the unobservable characteristics that shaped migrant's earnings.

This study is also related to the literature strand on the costs of international migration. Ortega and Peri (2009) provided estimates of the effects of immigration policy tightening on the magnitude of migration flows in 14 OECD countries over the period 1980 to 2005. They confirmed that, while South-North migration flows increase as a function of the per capita income gap between origin and destination, they decrease significantly when destination countries adopt stricter immigration laws. McKenzie and Rapoport (2010) highlighted the importance of network in shaping migrations costs. They found a positive selection for Mexican migrants leaving from communities with weak migrant networks, but negative self-selection for Mexican migrants from communities with stronger networks. In its attention to self-selection in unobservables, this work is also related to the large literature on selection-correction methods (Heckman, 1979; Lee, 1983; Dubin and McFadden, 1984; Dahl, 2002; Bourguignon et al., 2007). Finally, this paper also contributes to the literature analyzing the Venezuelan migration episode, which has so far only addressed the effect of the arrival of Venezuelans on the destination country (Bahar et al., 2020; Olivieri et al., 2022; Doocy et al., 2019; Namen et al., 2019; Caruso et al., 2021).

The remainder of the paper is organized as follows: Section 2 briefly overviews the salient features and the economic determinants of the Venezuelan exodus. In Section 3, I present the different sources of data and the relevant descriptive statistics. Section 4 deals with the empirical approach. Section 5, is devoted to the outline of the results, by presenting the main estimates regarding the coefficients of the main variable of interest. In conclusion, I provide, in Section 6, a discussion about the main findings, along with their potential policy implications.

2 The Venezuelan crisis

In this section, I discuss the nature of the Venezuelan migration crisis, which has been mainly directed to Colombia, Peru, Chile, and the United States. Venezuela has been facing the worst socioeconomic and political crisis ever experienced by any Latin American country. The beginning of this crisis dates back to the election of Hugo Chavez as president in 1998. During his presidency, Chavez introduced a socialist regime that included constitutional amendments, expropriations of land, implementation of populist social programs, nationalization, and restrictions on private companies. Nicolas Maduro, elected president of Venezuela in 2013, continued along the same lines as his predecessor, worsening the country's economic and social crisis. Populist policies, unsustainable public debt, low oil prices, and excessive and rigorous controls and regulations on the private sector have led to a deep economic recession. In early 2018 the Venezuelan government "essentially stopped" producing inflation estimates. However, according to Hanke (2018), by the end of 2018, the country was experiencing a hyperinflation rate of 33,151%.

The financial and economic crisis has led to a harsh humanitarian crisis, prompting an unprecedented wave of international emigration, with nearly 5.2 million Venezuelans leaving over the 2013-2019 period, from a country with a total population of 28.9 million [1]. Crucially for my analysis, Venezuelans moved essentially to four main destination countries, which absorbed 70 percent of the Venezuelan migration driven by the economic crisis. The most common destination was Colombia (1.8 million migrants). Peru welcomed 830,000 migrants, whereas Chile and the United States hosted 455,000 and 352,000 Venezuelans, respectively. Figure 2 plots the distribution of the Venezuelan migrants in the GEIH (2018), ACS (2018), ENPOVE (2018), and CASEN (2018) by year's quarter of arrival.

Interestingly, although this migration episode is highly concentrated in a few years and the timing is similar in all destinations, the scale differs substantially. Indeed, from 2013 to early 2017 about 130,000 Venezuelans had migrated to the United States, while only 85,000 had emigrated to Chile and 70,000 to Peru. In comparison, Venezuelan inflow to Colombia was three times larger over the same period. Specifically, Colombia had already welcomed 310,000 Venezuelan migrants. The sudden economic crisis and subsequent surge in Venezuelan emigration provide a clean exogenous shock that rules out the possibility that education may be endogenous to the prospect of emigration.

[1] According to the last Population Census in 2011

3 Data

3.1 Data sources

The analysis requires individual-level data for the Venezuelan population residing in the origin country and in the four main receiving countries over the relevant period. Specifically, the full dataset was created by combining five different data sources. For Venezuela, I use the 2018 Encuesta Nacional de Condiciones de Vida 2018 (hereafter ENCOVI). The survey was carried out by the Universidad Catolica Andres Bello de Caracas in 2018. It is represented by design of the Venezuelan population and provides information about 21,382 individuals, divided into 5,950 households across 22 states [2]. For Colombia, I rely on the Household National Survey (or GEIH in the Spanish acronym [3]), conducted at the end of 2018 on a sample of 14,128 Venezuelan migrants. The investigation concerns both the legal foreign population and the undocumented Venezuelan population. For Peru I use the 2018 ENPOVE (Instituto Nacional de Estadística e Informática, "Encuesta Dirigida a la Población Venezolana que Reside en el País 2018"). The survey, performed at the end of 2018, collects information on 9.847 Venezuelan migrants residing in Peru, which is the second-largest receiving country. It was carried out by the Peruvian National Institute for Statistics (INEI) between November and December 2018. It is represented by design of the Venezuelan population residing in Peru. In particular, it was conducted in the five largest cities in the country (Lima, Tumbes, Trujillo, Cusco, and Arequipa), where 85 percent of Venezuelans reside. For the United States, I chose the 2018 American Community Survey (U.S. Census Bureau, American Community Survey (ACS), Public Use Microdata Sample (Pums), 2018). Its sample is approximately 2.5 percent of the resident population in the US. For Chile I rely on the National Socioeconomic Survey 2018 (or Casen in the Spanish Acronym), conducted by the Ministry of Social Development. It provides information on 1.2 percent of the total population living in Chile.

Finally, I merged the five datasets, appropriately weighted. The full dataset contains information on age, employment status, occupation, and labor earnings. According to the existing literature on economic migration, I opted to restrict the sample to the working-age population (18 to 60). My final sample consists of 34,218 observations provided by 12,335 stayers, 12,983 Venezuelans settled in Colombia, 7,555 migrants in Peru, 580 migrants in

[2] The sample does not include Amazonas and Dependencias Federales

[3] Departamento Administrativo Nacional de Estadística, Gran Encuesta Integrada de Hogares - GEIH - 2018)

Chile, and 765 migrants in the United States.

The timing of the surveys described above fits well with the need to focus on Venezuelans who left their home country driven by Venezuela's sudden economic collapse. However, it would have been ideal to obtain data from more recent surveys, as the intense flow of migration, despite attenuating from 2018, continued into 2019 [4].

3.2 Descriptive statistics

Table 1 reports some basic descriptive statistics on stayers and migrants to the four destination countries. It shows that Venezuelans who emigrated to the United States are predominantly women (53 percent), and the majority of Venezuelans who moved to Peru are men (53 percent) In Colombia and Chile the sample is gender balanced. In line with the literature on migration selection, emigrants are younger than stayers, regardless of gender. The average age of the Venezuelans that chose not to leave the country was 37 at the time of the survey [5]. Venezuelan migrants residing in the US are the oldest (35 years old), whereas those in Colombia are the youngest (30 years old). For both genders, migrants are more educated than non-migrants. 19 percent of stayers have a university degree. Venezuelans who moved to Chile are the most educated. 65 percents of them are college graduates. The Venezuelans living in Peru are the least educated (27 percent of them have a university degree). The share of college graduates among Venezuelan women is higher than among men (64% in the US, 70% in Chile, 33% in Peru, and 32% in Colombia).

Turning to the earnings data, several features deserve careful consideration. First, I note a remarkable migration premium. Despite being younger, migrants earned much higher annual incomes than non-migrants. Additionally, Venezuelans in our sample are short-term migrants (they left the origin country in the last five years). This implies that in some years I may observe a much higher migration premium. Second, the data reveals a large and fairly constant female penalty in earnings. Although women are better educated, across the five locations, they earn at least 20 percent less than men. Third, university premiums differ greatly depending on the destination country. Peru shows the lowest return to education. On average Venezuelan migrants who settled in Peru and having a degree earn only 6% more than migrants with no university degree (3883$ compared to 3661$). Colombia has

[4] In any case, by the end of 2018, already more than 3 million Venezuelans had left the country. Note that Venezuelans included in my sample are short-term migrants, namely migrated within the last 5 years.

[5] Some World Bank estimates show a constant increase in the average age since the beginning of the crisis, due to a lower survival rate at birth, a lower birth rate and migration of young people.

much higher rates of return to education than Peru: the degree on average ensures 29% higher income. The US and Chile show the highest return to education: a college-graduated migrant earns 53% percent more than a non-college graduate.

The premium differences across destinations are not due to the length of migrants' stay. The clean "push shock" due to the sudden economic crisis led to mass migration in the short term. Most of the migrants in my sample migrated in 2017 and 2018. These descriptive statistics regarding earnings and education rates are in line with the results obtained (Grogger and Hanson, 2011), which state that the absolute differences in earnings between college and non-college individuals are higher in high-income countries. Although descriptive statistics suggest that migrants have the potential to obtain a large gain by moving from their home country (which is not surprising given the economic situation in Venezuela), it is unclear to what extent this is due to their observable and unobservable characteristics.

4 Empirical approach

As discussed in Section 2, the Venezuelan exodus provides a unique context for studying the self-selection of migrants, comparing migration to developing countries (South-South migration) and migration to developed countries (South-North migration). To assess migrants' self-selection, and to evaluate the role of expected earnings and migration costs in shaping the individual migration decision, I follow three distinct steps.

4.1 Self-selection with respect to education

First, I estimate a series of multivariate regression models related to how migrants differ in terms of observable variables from the population of Venezuelans who decided not to leave the country. I restrict the analysis to working-age Venezuelans (aged 18 to 60) who reported earnings other than zero, in order to focus on those most likely to have the decision-making power to migrate and to have completed the education path. The main variables of interest are age, gender, and education level. To facilitate interpretation, I opted to estimate linear probability models, although probit and logistic regression models have returned similar patterns. In addition, I estimate most models separately by gender.

The model of self-selection takes the form:

$$Migrant_{ij} = \alpha + \beta_1 X_i + \epsilon_i \tag{1}$$

where $Migrants_{ij}$ take value 1 if individual i migrated to country j and 0 otherwise. X_i is a vector of demographic variables, including age, age squared, and an education dummy that takes value 1 if the individual has a university degree.

4.2 Self-selection with respect to predicted earnings

In this subsection, I use the information on the annual earnings of Venezuelans to assess the self-selection with respect to predicted outcomes. As the migrants' surveys did not collect information on how much Venezuelans earned in their home countries before they left and, of course, each individual is only observed in one single location, the first step involves predicting individual-level earnings for all alternative locations, according to their individual observable characteristics. The availability of data on migrants' earnings in destination countries is of great value, especially for South-South migration episodes. Indeed, most empirical studies have had to rely on extrapolations from income figures for the general population. However, as shown by McKenzie and Rapoport (2010) in the context of Tongan migrants to New Zealand, migrants are likely to differ from non-migrants in several unobservable aspects. Even in the context of interest, there might be some unobserved characteristics that push people to move to a specific destination country and systematically bias their wage draws. For example, suppose that equally educated Venezuelans moving to the United States are less risk-averse people than Venezuelans migrating to Peru or Venezuelans who chose to stay in their home country. At the same time, they will be more likely to be hired in more risky and better-paid jobs. Therefore, unobserved heterogeneity in the propensity to migrate affecting also earnings would create a selection bias in the wage equation.

To address this concern, I apply the self-selection correction method proposed by Dahl (2002) [6]. First, I divided the population into 18 mutually exclusive cells, defined by gender, education (college versus high school versus non-educated), and age (three age groups). For each of these cells, I calculate the proportion of individuals who actually stays and works for

[6] The method was also used by Bertoli et al. (2013). In a South-North migration episode (from Ecuador to the US and Spain) they found that, including the Dahl parameter, the estimates of the expected earnings do not change much.

a wage in Venezuela ($\hat{p}_{i1}$), migrates to the US, and work for a wage there ($\hat{p}_{i2}$), to Colombia ($\hat{p}_{i3}$), to Peru ($\hat{p}_{i4}$), and migrates to Chile and works for a wage there ($\hat{p}_{i5}$). These terms represent the predicted probability that an individual belonging to a particular cell chooses to work in the respective location.

Next, I estimate the Mincer regressions by adding the Dahl's correction Polynomial $f_j(\hat{p}_{i,j}, \hat{p}_{i,j'})$. As suggested by (Bertoli et al., 2013), I include it as a second-order polynomial in the retention probability for stayers and a second-order polynomial in the retention and first-best probability for migrants plus an interaction term. This polynomial term aims to correct that migrant workers who move to a particular destination have a higher unobserved propensity to migrate and work there, which may also affect their earnings realization. The assumption behind Dahl's method is that the unobserved heterogeneity within cells is relatively small. Last, I predict log earnings $\hat{w}_{ij}$ in all five locations for all individuals in the sample.

Once I obtain selection-corrected individual wages, I use them to assess the self-selection of migrants with respect to predicted incomes. To this end, I apply the method suggested by Borjas et al. (2019), who showed that the same conditions that result in positive or negative selection in terms of expected earnings also imply a stochastic dominance relationship between the counterfactual earnings distribution of migrants in the origin country and the earnings distributions of stayers.

4.3 Location Choice

To evaluate the role of expected earnings and migration costs in shaping the individual migration decision, I estimate a discrete choice migration model, by using the predicted counterfactual earnings for all alternative locations.

I can summarize the model by:

$$U_{i,j} = \alpha ln(\hat{w})_{i,j} + x_i\beta_j + \upsilon_{i,j} \tag{2}$$

for each individual i in locations $j = 1, 2, 3, 4, 5$. In practice, I estimate a mixed logit discrete-choice model as main specification and a conditional logit model as robustness checks. The main explanatory variable for the probability of choosing a specific location is my estimate of predicted self-selection corrected log earnings at that location ($\hat{w}_{i,j}$). I

also include individual controls used in the previous steps: age, age squared, education, and gender (x_i) and country-specific intercepts to account for differences in policies, institutions, culture, and cost of living across the four destination countries. Although individual controls have already been used in earnings prediction, I include them in the decision model to capture individual differences in migration costs depending on the destination chosen. For example, women may face different migration costs depending on the family reunification policies implemented by the destination country.

5 Results and discussion

This section presents three sets of results. I first show self-selection of migrants according to education and other individual observable characteristics. Second, I present results regarding self-selection of migrants with respect to predicted earnings. Last, I show results for the discrete-choice migration model, focusing on the estimates of the expected earnings coefficient and the heterogeneity of the migration costs by gender and destination country.

5.1 Self-selection with respect to education

The descriptive statistics reported in section 3.2 suggest that Venezuelan migrants living in Chile and in the United States were substantially more educated than the Venezuelan individuals residing in Colombia, Peru, or in the origin country. In this subsection I provide more rigorous results, controlling for a set of individual observable characteristics, such as age and gender.

Table 2 presents evidence for the full migrant sample and separately for the four destination countries. I also estimate each specification for males and females. Outcome variables are as follows: *migrant to any country* is the outcome variable in Columns 1 and 2; migrants who moved to Colombia in Columns 3 and 4; Venezuelan migrants in Peru in Columns 5 and 6; migrants in Chile in Columns 7 and 8, and migrants in the US in Columns 9 and 10. In each analysis, the control group of stayers consists of ENCOVI respondents. I also present two-sided tests on whether the college dummy coefficient is statistically different across regressions for each country. In columns 1 and 2 of Table 2 I find that, for both males and females, educated people are significantly more likely to migrate, which means that migrants are positively selected in education with respect to stayers. However, as sug-

gested by the two-sided tests of coefficient equality across regressions, the effect of having a college degree on the probability of migrating is different depending on the country of destination and gender. Indeed, Table 2 shows a weaker positive selection for Venezuelan females who moved to Colombia and for Venezuelan males who moved to Peru (respectively Columns 4 and 5 of Table 2). This might be due to the small wage premium for females in Colombia and males in Peru (Table 1). Stronger positive self-selection of people moving to more developed countries is in line with Borjas (1990), as the US and Chile have much wider income differences than Venezuela (Grogger and Hanson, 2011).

This table clearly suggests that the skills composition of the migration flow differs greatly according to the income level of the destination country. Indeed, the skill composition of the South-North migration flow (from Venezuela to Chile or to the United States) appears much better than the skill composition of the South-South migration flow (from Venezuela to Peru or Colombia) [7].

5.2 Self-selection with respect to predicted earnings

Table 3 shows the results of the Mincer model estimates. The odd-numbered data columns report the results of the Mincer regressions without the addition of the Dahl term. First, I find a very large female wages penalty in the five countries, although larger in the United States (51%) than in Chile (17.2%), Peru (17%), and Colombia (32%). Then, it's worth noting that the estimates imply a large college premium in the origin country (37.5%) and Chile (34.7%). The college premium is about the same in Colombia and the US, 25 and 22.6 percent, respectively. In Peru having a degree ensures the lowest premium among the countries under analysis. This is in line with the low skill composition of migrants in Peru.

The even data columns of Table 3 show the Dahl earnings regression estimates, which correct for self-selection in unobservables. First, I observe that estimates are lower than the previous specification for all the locations. In addition, for the cases of Venezuela, Peru, and the US we reject the null hypothesis of joint zero values of the correction Dahl parameters. It suggests that unobserved individual characteristics play an important role in predicting the wage. Specifically, when adding the correction Dahl terms, the estimated college premium in Venezuela decreases to 16.9%. For the case of Peru, it decreases to

[7] Two probit models are also estimated, with which I measure the selection and sorting of migrants (In appendix, Table ??). The results confirm the positive selection of migrants compared to stayers and the positive sorting of Venezuelans living in Chile and the US compared to Venezuelans living in Peru and Colombia.

3.9%, consistent with the expectation that the less educated migrants move to countries with the lowest college premium. In the case of the United States, the coefficient drops to 0.152, and it is no longer significant, suggesting that unobservable characteristics in the United States play a crucial role. One potential explanation is that there may be greater skill downgrading in the United States than in other destination countries, and thus there may be more important factors than formal education in predicting salaries. Some of these factors could be English proficiency or the network at destination.

Figure 3 compares cumulative distribution functions of migrants' predicted income (in the origin country) with stayers' CDFs. This figure clearly suggests that migrants in US and Chile are positively selected in terms of predicted incomes. Migrants to Colombia and Peru, on the other hand, appear to be negatively selected over the stayers. Figures 4 and 5 show the results based on gender. For male migrants, the results do not change much from the previous figure. The female Figure 5, instead, shows that only women who migrated to the United States actually differ from stayers.

The results regarding South-South migration self-selection in terms of predicted income seem to be contradictory to those on self-selection in terms of observable characteristics. As a matter of fact, migrants who moved to Peru or Colombia are better educated than stayers, but they are negatively selected in terms of predicted income. This leads to two considerations. First, unobservable characteristics play a key role in identifying the level of productivity of labor migrants. Second, education level still plays a role in the migration decision, regardless of the migrant's productivity. Namely, a high education could decrease migration costs by increasing the likelihood that a college graduate could obtain a visa and/or passport. In the next subsection, I test this hypothesis by estimating the effect of education on migration decisions while controlling for the migration premium.

5.3 Location choice: earnings and costs of migration

Table 4 shows the estimates for five specifications. Column 1 presents my main results. The earning coefficient α is positive and highly significant (0.502), suggesting that higher expected earnings in a particular country increase the probability to move there. The estimates are robust across the conditional logit model specifications, in which the effects of the individual characteristics do not vary across alternatives. Again, the coefficient of expected earning is positive and highly significant. The three specifications clearly suggest that earnings at destination are a robust determinant of migration choices.

I turn now to the interpretation of the coefficients of individual controls and country intercepts. First, I note that the coefficients associated with the education variable are still positive, despite the fact that I am controlling for earnings. This confirms that people with college degrees, conditional on expected earnings, may face lower migration costs (Column 1 in Table 4). One potential explanation may lie in welcoming migration policies for college graduate migrants or better networking at destination. Second, being female is associated with a higher probability to migrate to the US, Colombia, and Peru. Finally, it is worth noting that the coefficients of the country intercepts are high and strongly significant. Moreover, we can see that they differ greatly across destination countries. The fact that individual characteristics are strongly significant despite controlling for expected earnings at destination suggests that migration costs may play a key role in understanding this migration flow.

5.3.1 Costs of Migration

Understanding the nature of the migration costs faced by Venezuelan migrants is key to the analysis of this specific migration episode and to properly implementing migration policies in destination countries and the country of origin. Although Venezuela has been facing a very deep economic crisis that has brought many Venezuelans close to the poverty line, large numbers of them have not yet left the country. This implies that migration costs are very high compared to the difference in earnings. Of course, the costs go far beyond the cost of physically moving from Venezuela to the destination country. There may be indirect costs associated with being away from family and friends but also costs of overcoming barriers to migration. For example, for legal migrants, these can be the months of waiting for a visa. For irregular migrants, on the other hand, may be related to the risk of being caught by the police or taking very dangerous migration routes.

Furthermore, the question of whether restrictive and/or receptive migration policies are decisive remains an open and critical policy question. In this respect, the effect of restrictive migration policies implemented by farther away countries (without the possibility of overland migration) may have been a barrier to migration, given also the greater ease of border control. In the case of the United States, March 2017 was a key turning point for Venezuelan migration flow (Figure 2). As of July 2017, the number of Venezuelans residing in the United States had increased by 167,105 since 2010, bringing the total to 351,144 migrants (IOM, April 2018). Since the Trump administration, the United States

has been virtually closed to Venezuelans. US officials could detain Venezuelans arriving on tourist visas if inspectors suspected they intended to stay rather than visit. As a result, the number of tourist visas issued to Venezuelans dropped to 47,942 after reaching 239,772 in 2015. Moreover, as of 2017, the United States increased the cost per visa application from 30 (2015) to 160 US Dollars. The price was prohibitive, as I showed that the average salary of a Venezuelan in 2018 was about 250 US Dollars per year. The combination of these migration policies, together with the risk of migrating illegally, may have been the cause of this large increase in migration costs to the United States.

On the other hand, the effect of migration policies on the neighboring countries is still unclear. As shown in the Figure, there are some instances where migration policies have worked in their intent (passport request in Peru in August 2018), but I also find events where receptive policies are followed by decreased inflows (Colombia, February 2018) and restrictive policies by increased inflows (Peru, March 2019).

In conclusion, it does not seem that policy changes had a major and constant influence on migration flow. This may probably also be due to the impossibility of border control and irregular migration supported by organized crime. Therefore, I believe the perception that restrictive policies succeed in limiting migrant inflows needs to be revised, at least in the context of South-South migration and neighboring countries. Beyond border policies, more attention needs to be paid to interventions that successfully affect pull and push factors.

6 Conclusion

In this paper, I examine migrants' self-selection in the context of the Venezuelan exodus by combining five different data sources from the origin country and the main destination countries.

The results show that, for both males and females, migrants are positively selected in terms of education. Moreover, South-North migrants appear positively sorted compared to South-South migrants. I also find that controlling for self-selection in unobservables, migrants located in Chile and the US are positively selected compared to stayers in terms of predicted income. South-South migrants, instead, appear negatively selected compared to Venezuelans who chose not to leave the country. Such a result implies that unobservable characteristics play a key role in identifying the level of productivity of migrants. Finally, the results obtained from the estimation of the discrete-choice migration model suggest

that even controlling for expected earnings in the destination country, individuals may face different migration costs depending on their characteristics (age, gender, and education). In particular, being female is associated with a higher likelihood of migrating to the United States, Colombia, and Peru, indicating that many people migrate thanks to family reunification policies. In addition, having a degree seems to greatly lower migration costs, especially in South-North migration.

These findings extend the literature on migrant self-selection, bringing empirical evidence in a context where South-South migration outweighs South-North migration. Although observable characteristics explain much of the variance in wages, unobservable characteristics also play an important role in estimating the migrant's level of productivity. In addition, distinguishing and disaggregating the importance of each migration cost faced by Venezuelan migrants is crucial for the analysis of this specific migration episode and for properly implementing migration policies in both destination countries and the country of origin. Although this current work illuminates the importance of self-selection and migration costs in shaping the migration decision, further studies focusing on South-South migration and effective policy interventions are highly recommended.

Figures and Tables

Table 1: Descriptive Statistics

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Venezuela		Colombia		Peru		Chile		U.S.	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Female	0.52	0.50	0.51	0.50	0.47	0.50	0.50	0.50	0.53	0.50
Age	37	12	30	9	31	9	32	9	35	10
College	0.19	0.39	0.28	0.45	0.27	0.44	0.65	0.48	0.57	0.50
Labor income, 2018 USD	249	539	2,122	1,328	3,720	1,605	8,896	7,054	25,475	26,844
Non-college graduate	241	523	1,956	1,074	3,661	1,517	6,533	3,890	19,516	15,136
College graduate	276	594	2,515	1,731	3,883	1,818	10,001	7,887	29,789	32,152
College premium	0.15		0.29		0.06		0.53		0.53	
Female										
Age	36.80	12.44	30.17	9.24	31.33	9.67	33.41	9.70	35.75	9.81
College graduate	0.23	0.42	0.32	0.47	0.33	0.47	0.70	0.46	0.64	0.48
Labor income, 2018 USD	216	527	1,881	1,317	3,399	1,555	8,233	6,028	20,009	23,641
Male										
Age	36.82	12.42	30.21	9.00	30.83	8.50	31.59	8.43	34.90	10.14
College graduate	0.15	0.36	0.24	0.43	0.21	0.41	0.61	0.49	0.50	0.50
Labor income, 2018 USD	268	545	2,279	1,311	3,955	1,602	9,412	7,733	30,210	28,541
N		12,235		12,983		7,555		580		765

Source: Author's elaboration on ENCOVI 2018, GEIH 2018, ENPOVE 2018, ACS 2018 and CASEN 2018.

Table 2: Self-selection of Venezuelan migrants, adults aged 18-60

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	All Migrants		Colombia		Peru		Chile		US	
	Males	Females	Males	Females	Males	Females	Males	Females	Males	Females
College graduate	0.117***	0.117***	0.124***A	0.096***A	0.088***A	0.117***A	0.142***A	0.104***A	0.131***A	0.116***A
	(0.008)	(0.008)	(0.010)	(0.009)	(0.013)	(0.011)	(0.011)	(0.008)	(0.012)	(0.009)
Age	0.027***	0.015***	0.015***	0.009***	0.028***	0.006**	0.002*	0.001	0.005***	0.007***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)	(0.001)	(0.001)	(0.002)	(0.002)
Age sq. * 1000	-0.541***	-0.354***	-0.379***	-0.300***	-0.520***	-0.205***	-0.051***	-0.032**	-0.077***	-0.098***
	(0.029)	(0.029)	(0.032)	(0.031)	(0.035)	(0.035)	(0.015)	(0.015)	(0.020)	(0.019) height
N	16,876	17,039	12,189	12,826	9,825	9,769	6,084	6,535	6,154	6,650
R^2	0.108	0.091	0.106	0.099	0.091	0.066	0.075	0.053	0.050	0.053

Notes: Robust standard errors in parentheses and they allow for State Level clustering. Asterisks denote statistical significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$ Notes: Robust standard errors in parentheses. Outcome variable, migrant, is equal to 1 for Venezuelans who left the origin country and 0 for respondents in Venezuela at the time of the ENCOVI survey. A Indicates a statistically significant difference in each pair of estimates on the effects of education between refugees and irregular migrants at p ¡ .05. ***p ¡ .01, **p ¡ .05, *p ¡ .1.

Table 3: Determinants of Labor Earnings

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Venezuela		Colombia		Peru		Chile		U.S.	
	Mincer	Dahl	Mincer	Dahl	Mincer	Dahl	Mincer	Dahl	Mincer	Dahl
College graduate	0.375***	0.169*	0.251***	0.195***	0.070***	0.039**	0.347***	0.287*	0.226**	0.152
	(0.061)	(0.090)	(0.017)	(0.032)	(0.012)	(0.019)	(0.060)	(0.168)	(0.106)	(0.192)
Female	-0.343***	-0.379***	-0.316***	-0.327***	-0.175***	-0.194***	-0.172***	-0.160***	-0.513***	-0.498***
	(0.052)	(0.053)	(0.016)	(0.017)	(0.010)	(0.012)	(0.053)	(0.059)	(0.098)	(0.100)
Age	0.042***	0.061***	0.039***	0.045***	0.012***	0.015***	-0.003	-0.007	0.085**	0.113***
	(0.016)	(0.017)	(0.006)	(0.006)	(0.004)	(0.004)	(0.022)	(0.023)	(0.035)	(0.042)
Age squared*1,000	-0.448**	-0.527***	-0.553***	-0.593***	-0.189***	-0.221***	0.106	0.130	-0.914*	-1.040*
	(0.199)	(0.201)	(0.082)	(0.083)	(0.054)	(0.055)	(0.307)	(0.312)	(0.475)	(0.551)
N	5,650	5,650	8,045	8,045	6,670	6,670	427	427	543	543
R^2	0.014	0.016	0.068	0.071	0.048	0.052	0.100	0.101	0.093	0.100
P-Value, F-test										
Dahl polyn.		0.00184		0.438		0.0513		0.510		0.0632

Notes: Standard errors in brackets. The Dahl polynomial includes second-order polynomial in cell probabilities. Sample restricted to working-age population (aged 18 to 60). Asterisks denote statistical significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Source: Author's elaboration on ENCOVI 2018, GEIH 2018, ENPOVE 2018, ACS 2018 and CASEN 2018.

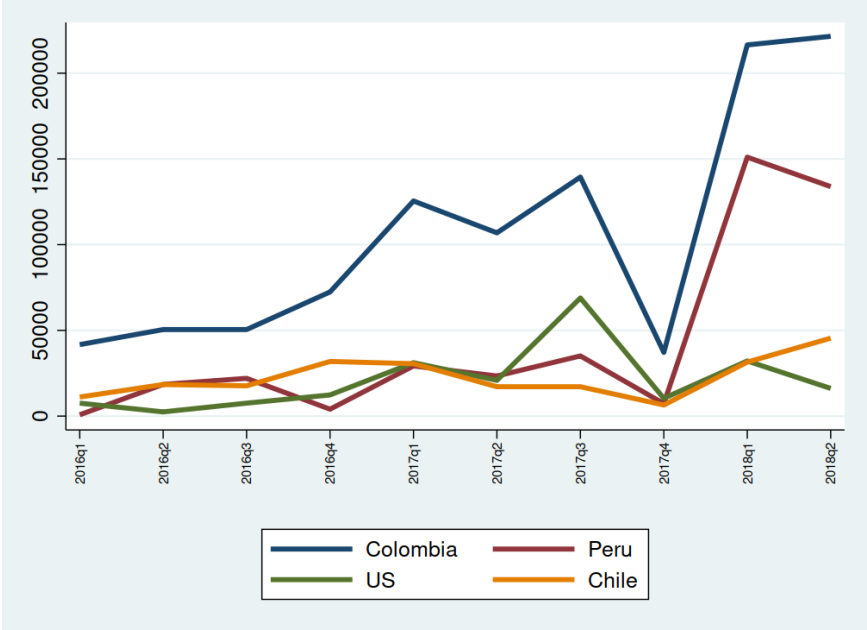
Table 4: Location Choice Model

	(1)	(2)	(3)
Specification	Mixed	Conditional	Conditional
Earnings	Logarithm	Logarithm	Logarithm
Self-selection term	Dahl	Dahl	Mincer
Earnings	0.502***	1.832***	2.081***
	(0.152)	(0.684)	(0.668)
Colombia intercept	-0.084	-6.069**	-6.942**
	(0.631)	(2.786)	(2.838)
College graduate	0.493***	-	-
	(0.046)		
Age	-0.021	-	-
	(0.013)		
Female	0.105***	-	-
	(0.038)		
Age sq*1000	-0.001***	-	-
	(0.000)		
Peru intercept	-1.587**	-7.549*	-8.588**
	(0.804)	(4.037)	(4.135)
College graduate	0.396***	-	-
	(0.063)		
Age	0.037**	-	-
	(0.015)		
Female	0.168***	-	-
	(0.047)		
Age sq*1000	-0.001***	-	-
	(0.000)		
Chile intercept	-6.284***	-11.473***	-12.693***
	(1.139)	(4.068)	(4.031)
College graduate	2.046***	-	-
	(0.112)		
Age	0.097**	-	-
	(0.044)		
Female	-0.072	-	-
	(0.109)		
Age sq*1000	-0.002***	-	-
	(0.001)		
USA intercept	-7.016***	-12.713**	-14.043***
	(1.033)	(5.159)	(5.009)
College graduate	1.617***	-	-
	(0.098)		
Age	0.090***	-	-
	(0.034)		
Female	0.237**	-	-
	(0.094)		
Age sq*1000	-0.002***	-	-
	(0.000)		
Number of cases	21,417	21,417	21,417
Pseudo R squared		0.230	0.230
Log-likelihood	-25374	-26535.84	-26539.54

Notes: Standard errors in brackets. I normalized the coefficients for Venezuela of the variable which do not vary across destinations to zero. Asterisks denote statistical significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

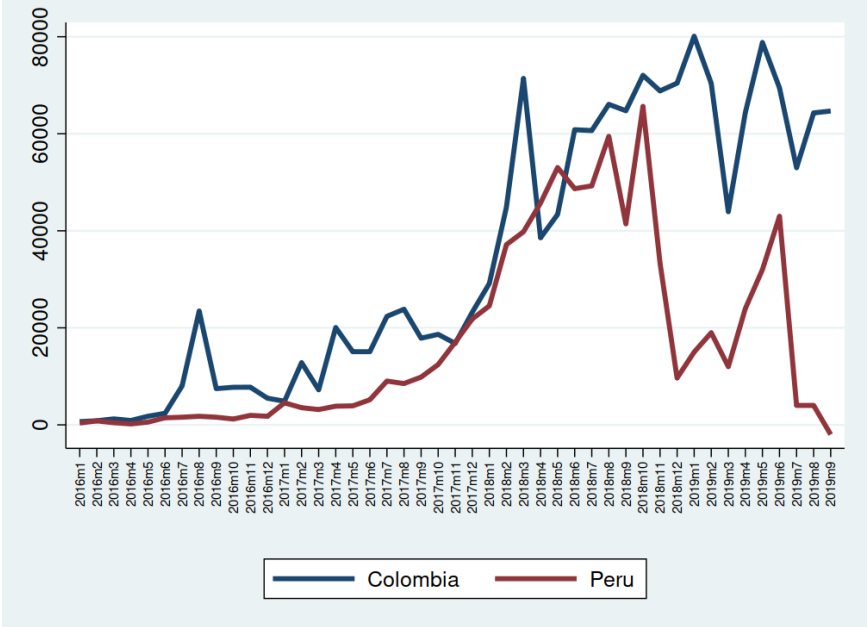
Source: Author's elaboration on ENCOVI 2018, GEIH 2018, ENPOVE 2018, ACS 2018 and CASEN 2018.

Figure 1: Arrivals of Venezuelans to Colombia, Peru, U.S. and Chile (2016-2018)



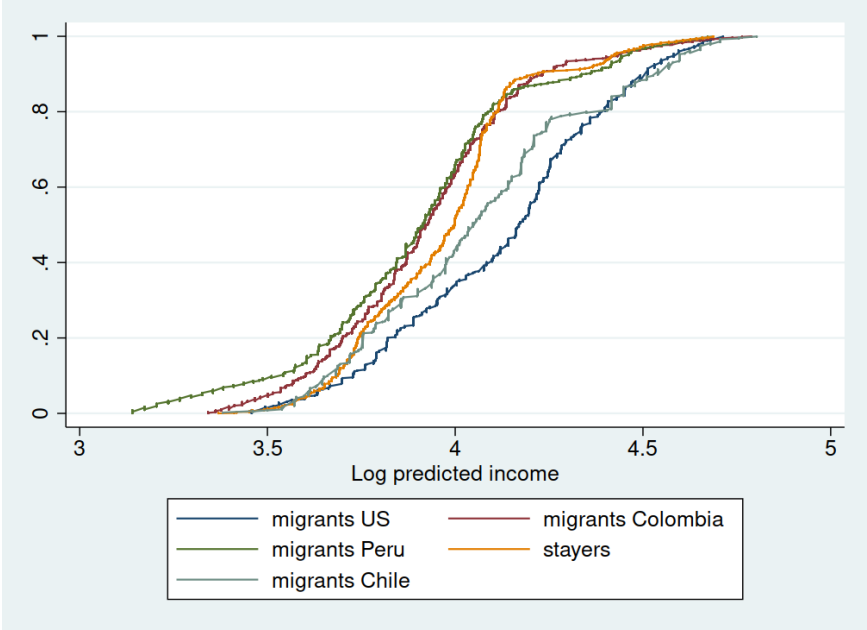
Notes: Author's calculations using GEIH (2018), ENCOVI (2018), ACS (2018), ENPOVE (2018) and CASEN (2018)

Figure 2: Monthly arrivals of Venezuelans to Colombia and Peru (2016-2019)



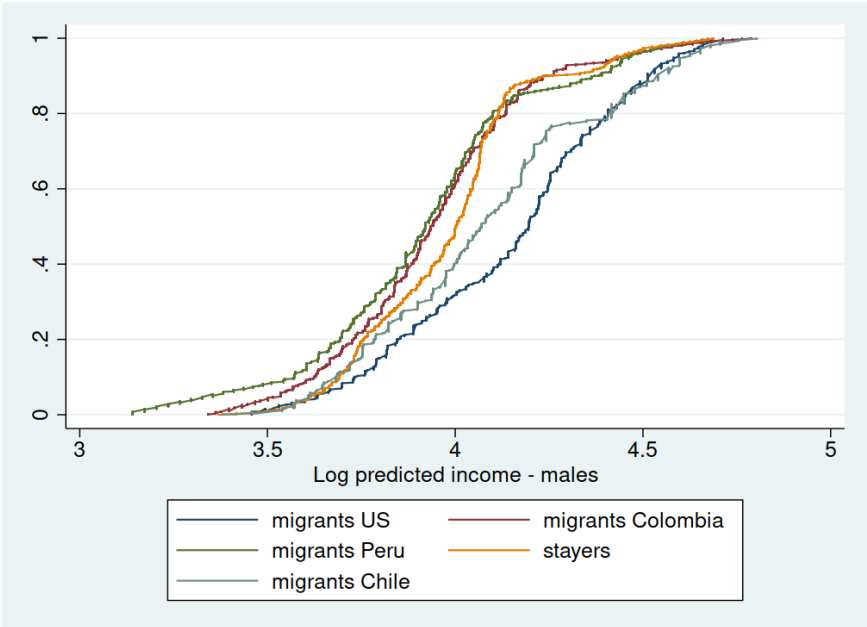
Notes: Oficina de Migraciones (Peru) and Ministerio de Relaciones Exteriores (Colombia)

Figure 3: Cumulative distribution functions for Venezuelan migrants' and stayers' predicted income in the origin country



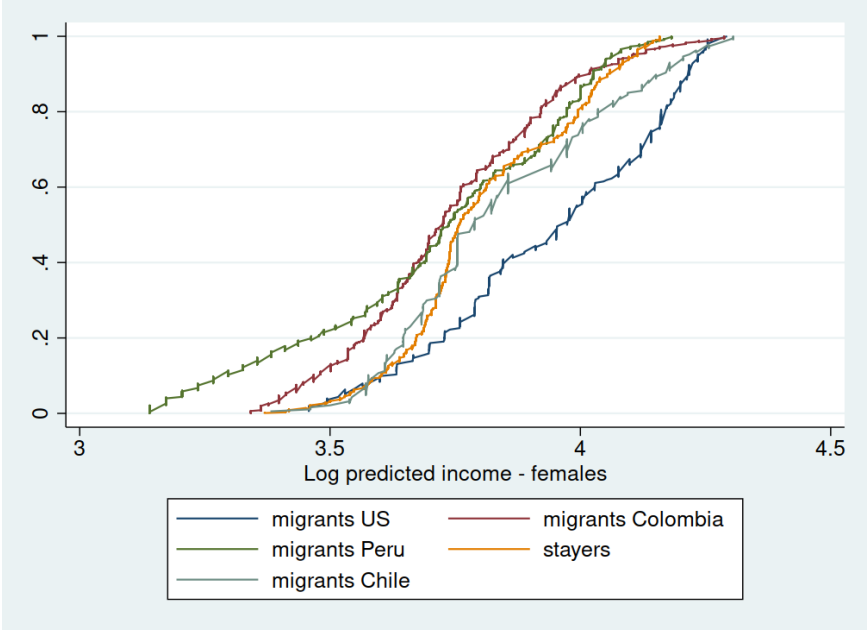
Notes: These charts present cumulative distribution functions of Venezuelan migrants' and stayers' predicted income in the origin country.
Sources: ENCOVI 2018, ACS 2018, ENPOVE 2018, GEIH 2018 and CASEN 2018.

Figure 4: Cumulative distribution functions of Venezuelan migrants' and stayers' predicted income in the origin country (Male sample)



Notes: These charts present cumulative distribution functions of Venezuelan migrants' and stayers' predicted income in the origin country (Male sample).
Sources: ENCOVI 2018, ACS 2018, ENPOVE 2018, GEIH 2018 and CASEN 2018.

Figure 5: Cumulative distribution functions for Venezuelan migrants' and stayers' predicted income in the origin country (Female sample)



Notes: These charts present cumulative distribution functions for Venezuelan migrants' and stayers' predicted income in the origin country (Female sample).
Sources: ENCOVI 2018, ACS 2018, ENPOVE 2018, GEIH 2018, and CASEN 2018.

Appendix

Table 5: Selection and Sorting Regressions, Probit Model

	(1)	(2)	(3)	(4)	(5)
	Selection in education			Sorting in Education	
	Full Sample	Male	Female	Peru – Colombia	Chile - U.S.
Colombia migrant dummy variable	0.307***	0.344***	0.275***	0.066***	-
	(0.019)	(0.028)	(0.025)	(0.020)	
Peru migrant dummy variable	0.238***	0.199***	0.282***	-	-
	(0.021)	(0.031)	(0.029)		
Chile migrant dummy variable	1.250***	1.277***	1.224***	-	-
	(0.056)	(0.078)	(0.081)		
U.S. migrant dummy variable	0.999***	0.984***	1.012***	-	-0.345***
	(0.049)	(0.070)	(0.068)		(0.075)
N	33,915	16,876	17,039	20,531	1,345

Notes: Standard errors in brackets. The dependent variable is having a college degree. Column 4 includes only migrants to Colombia and Peru. Column 5 includes only migrants to Chile and US. Asterisks denote statistical significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Source: Author's elaboration on ENCOVI 2018, GEIH 2018, ENPOVE 2018, ACS 2018 and CASEN 2018.

References

Bahar, Dany, Meagan Dooley, and Andrew Selee, "Venezuelan migration, crime, and misperceptions: A review of data from Colombia, Peru, and Chile," 2020.

Bertoli, Simone, J Fernández-Huertas Moraga, and Francesc Ortega, "Crossing the border: Self-selection, earnings and individual migration decisions," *Journal of Development Economics*, 2013, *101*, 75–91.

Borjas, George J, "Self-selection and the earnings of immigrants: reply," *The American Economic Review*, 1990, *80* (1), 305–308.

_ , Ilpo Kauppinen, and Panu Poutvaara, "Self-selection of emigrants: Theory and evidence on stochastic dominance in observable and unobservable characteristics," *The Economic Journal*, 2019, *129* (617), 143–171.

Bourguignon, François, Martin Fournier, and Marc Gurgand, "Selection bias corrections based on the multinomial logit model: Monte Carlo comparisons," *Journal of Economic surveys*, 2007, *21* (1), 174–205.

Bull, Benedicte and Antulio Rosales, "The crisis in Venezuela," *European Review of Latin American and Caribbean Studies/Revista Europea de Estudios Latinoamericanos y del Caribe*, 2020.

Caruso, German, Christian Gomez Canon, and Valerie Mueller, "Spillover effects of the Venezuelan crisis: migration impacts in Colombia," *Oxford Economic Papers*, 2021, *73* (2), 771–795.

Chiquiar, Daniel and Gordon H Hanson, "International migration, self-selection, and the distribution of wages: Evidence from Mexico and the United States," *Journal of political Economy*, 2005, *113* (2), 239–281.

Dahl, Gordon B, "Mobility and the return to education: Testing a Roy model with multiple markets," *Econometrica*, 2002, *70* (6), 2367–2420.

Doocy, Shannon, Kathleen R Page, Fernando De la Hoz, Paul Spiegel, and Chris Beyrer, "Venezuelan migration and the border health crisis in Colombia and Brazil," *Journal on Migration and Human Security*, 2019, *7* (3), 79–91.

Dubin, Jeffrey A and Daniel L McFadden, "An econometric analysis of residential electric appliance holdings and consumption," *Econometrica: Journal of the Econometric Society*, 1984, pp. 345–362.

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

Hanke, Steve, "IMF Produces Another Bogus Venezuela Inflation Forecast," https://www.forbes.com/sites/stevehanke/2018/07/31/imf-produces-another-bogus-venezuela-inflation-forecast/?sh=46a98f637dc8 2018. Last checked on 2022/07/21.

Heckman, James J, "Sample selection bias as a specification error," *Econometrica: Journal of the econometric society*, 1979, pp. 153–161.

Kaestner, Robert and Ofer Malamud, "Self-selection and international migration: New evidence from Mexico," *Review of Economics and Statistics*, 2014, *96* (1), 78–91.

Lee, Lung-Fei, "Generalized econometric models with selectivity," *Econometrica: Journal of the Econometric Society*, 1983, pp. 507–512.

McKenzie, David and Hillel Rapoport, "Self-selection patterns in Mexico-US migration: the role of migration networks," *the Review of Economics and Statistics*, 2010, *92* (4), 811–821.

Moraga, Jesús Fernández-Huertas, "New Evidence on Emigrant Selection," *The Review of Economics and Statistics*, 02 2011, *93* (1), 72–96.

Namen, Olga, Mounu Prem, Sandra Rozo, and Juan F Vargas, "Effects of Venezuelan migration on education outcomes in Colombia," *Research Prosals Inter-American Bank of Development*, 2019.

Olivieri, Sergio, Francesc Ortega, Ana Rivadeneira, and Eliana Carranza, "The Labour Market Effects of Venezuelan Migration in Ecuador," *The Journal of Development Studies*, 2022, *58* (4), 713–729.

Ortega, Francesc and Giovanni Peri, "The causes and effects of international mobility: Evidence from OECD countries 1980% 2005," *NBER Working Paper*, 2009, *14833.*

Rozo, Sandra V and Juan F Vargas, "Brothers or invaders? How crisis-driven migrants shape voting behavior," *Journal of Development Economics*, 2021, *150*, 102636.