# Understanding the Economic Impacts of Venezuelan Migrants and Refugees in Colombia

William Sims, FICSS UNHCR (sims@unhcr.org)

Introduction

The past 4 years have seen a marked increase in emigration from Venezuela to neighboring countries as a result of economic instability. A significant number of these migrants have arrived in Colombia, concentrated in departments near the Venezuelan border. The map below shows the departments with the highest percentage of Venezuelan refugees as a share of the population. Clearly visible in the graph, refugees are concentrated near the Venezuelan border, as well as the department of Cudinamarca near Bogota.
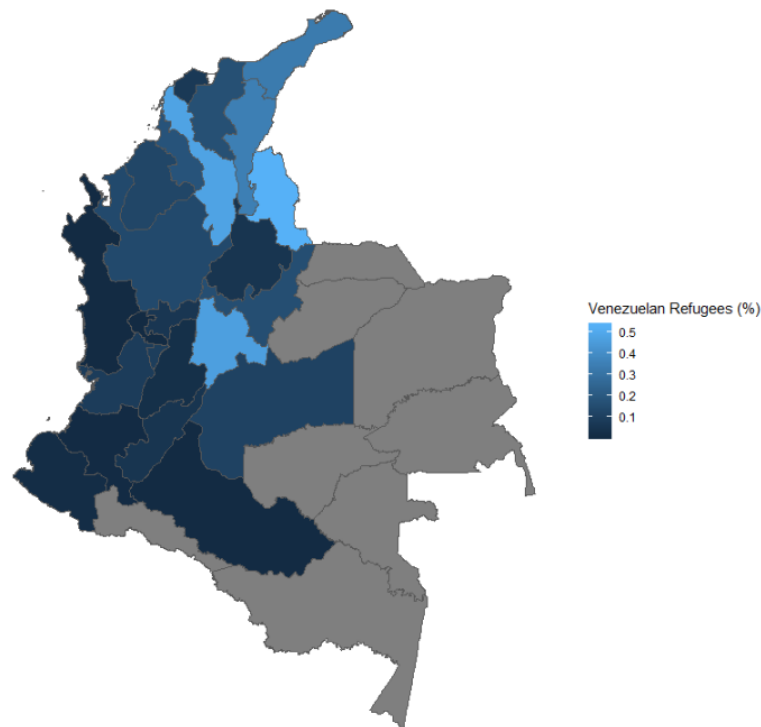
This study will focus on the economic well-being of native Colombians, as measured by employment, entrepreneurship, and income. The goal is to establish whether native Colombians living in departments who have received a significant influx of refugees and migrants are economically better or worse off as a result. There is some evidence to suggest that refugees and migrants can compete with locals for low-skill jobs, driving higher unemployment and lower wages.[1] Other research has found that proximity to refugees can drive higher rates of business ownership, as locals start businesses to serve the new migrant markets.[2] Although there has been previous work looking at the effects of internal displacement in Colombia, so far no research has examined the impact of migration in the context of the Venezuelan crisis.

---

[1] Verme, Schuettler (Forthcoming): *"The Impact of Forced Displacement on Host Communities: A Review of the Empirical Literature in Economics"*

[2] Loschmann, Bilgili, Siegel 2019: *"Considering the benefits of hosting refugees: evidence of refugee camps influencing local labour market activity and economic welfare in Rwanda"*

Data

      The analysis is based on Colombia's *Gran Encuesta Integrada de Hogares* (GEIH) household survey between 2013 and 2018. Much of the survey data is publicly available from Colombia's national statistics office (DANE), including all relevant household economic data, and individual characteristics like age, gender, and education level. Immigration status is established based on a non-public GEIH dataset that asks respondents about their place of birth, as well as where they were living 12 months and 5 years ago. The survey also asks about reason for migration for those who have moved within the last 12 months. The survey allows aggregation at the departmental level, but doesn't include eastern departments of the country that are primarily low population density jungle.

Definitions

      Economic

      The economic well-being of Colombians in this study is characterized by three key variables: employment, wages, and business ownership. Employment status is established through Question P6240 on the GEIH, which asks respondents what activity occupied a majority of their time in the past week. Responses include Working, Looking for Work, Studying, Housework, Disabled, and Other Activity. This allows analysis on two related yet distinct attributes. The employment rate captures only the share of respondents who responded that they spent a majority of their time working, thus not distinguishing between the various non-working alternatives. This would be a problem if there were endogenous differences in the rate of alternatives to work – for instance if students are concentrated in a particular department. On the other hand, the unemployment rate looks specifically at individuals who spent a majority of their time in the past week actively looking for work. This group explicitly does not include individuals who are not employed but have dropped out of the workforce entirely, perhaps due to discouragement with job prospects. It also excludes individuals who might be willing to work, but have been unable to find employment and so have returned to school or assumed a higher share of domestic labor. This analysis will thus consider both employment and unemployment in order to provide the most comprehensive picture of Colombians' employment prospects and labor market.

      Wages are defined as the stated response to a GEIH question on labor income. It doesn't include alternative sources of income such as pensions, rents, or one-time sales of goods. Thus while it doesn't provide a comprehensive picture of household income, it is the clearest indicator of pressure on wages due to competition in the labor market. In order to ensure consistency across the time period of the analysis, wages are adjusted annually for inflation relative to 2013.

      Finally, business ownership is derived from GEIH question P6765. The question asks all individuals who have reported that they are working what type of work they did during the past week. Possible responses include hourly or salaried work, work paid by the job, piecework, work on commission, catalog sales, self-employed, and owning a business in industry, commerce, services, or agriculture. Respondents in this last category are considered business owners for the purposes of this analysis. This excludes individuals like taxi drivers, plumbers,

and dayworkers who are independent in their work, but fall into the self-employed category. While the GEIH questionnaire is extensive, the economic analysis in this study will focus on these three indicators, leaving the other questions for future research.

Migration

The GEIH does not directly ask about migration status, so migration trends must be inferred from respondents' answers about their birthplace and previous residences. The GEIH asks where respondents were born, where they lived 5 years ago, and where they lived 12 months ago. All questions include an international option that has Venezuela available as a response.

For the purposes of this research, Venezuelan migrants are individuals who were born in Venezuela and were living in Venezuela 5 years and/or 12 months ago. Venezuelan refugees are Venezuelan migrants who answered that the primary motivation for their move in the last 12 months was due to threat or risk to their life or physical integrity not caused by armed conflict. This data is aggregated at the department level to establish the share of the surveyed population that is considered a Venezuelan migrant or refugee.

The analysis also defines Colombians as individuals who were born in Colombia. Colombia doesn't grant birthright citizenship, so this population is not synonymous with Colombian nationals, but the survey doesn't have a question on nationality so this is the closest definition possible. The definition used in the analysis also excludes individuals who were born in Colombia but were in another country 5 years or 12 months ago. These individuals have left Colombia for an unspecified amount of time, so it is impossible to determine whether they fit the profile of "native worker" that is the focus of this analysis.

Individuals who are of neither Colombian nor Venezuelan birth are also deliberately excluded from the analysis. Individuals in the dataset who have moved to Colombia within the past 5 years tend to have distinct education and income profiles from both native Colombians and Venezuelan migrants, and so they will be explicitly removed.

In total, there are 4,514,075 individuals in the data classified as Colombian, with an additional 14,858 who are Colombian and previously lived in another country, and another 19,958 Colombians who have returned to Colombia from Venezuela within the last 5 years. There are 24,627 Venezuelan Migrants, of whom 1,918 can be classified as refugees. There are also 7,701 Venezuelan-born respondents who have lived in Colombia for longer than 5 years, and 9,197 foreign-born non-Venezuelan immigrants. Finally, 96,326 respondents not answer the questions on nationality or migration at all, or did not respond adequately enough to permit a categorization.

General Characteristics

In addition to the economic and migration variables, there are additional independent control variables that can be used to normalize across departments. Gender, age, years of education, and urban/rural are all directly drawn from respondent responses. When aggregated at the department level, numerical variables like age and years of education are summarized by mean. Binary variables like gender and urban/rural are aggregated as proportions of responses.

Skill level is defined based upon the respondent's education level – following generally accepted definitions in the economic literature, individuals with a high school education or less ("Media 10o-13o") are considered low-skill workers. Workers with any post-high school education are considered high-skill. In the dataset, 22% of respondents are high skill, 73% are low skill, and 4% didn't answer. Formal work is defined as working individuals who have a verbal or written contract, representing 52% of respondents.

Methodology

The analysis is based on a time-series panel linear regression at the Departmental level. The identification strategy will rely on the unique nature of the current immigration influx. The number of Venezuelan immigrants to Colombia has increased dramatically over a relatively short time period – from just 0.1% of the GEIH surveyed population in 2013 to 1.3% in 2018. While most departments continue to have nearly nonexistent levels of Venezuelan immigration, border departments like La Guajira Norte de Santander have experienced levels of migrants that have reached over 5% as a share of the surveyed population. Migrant and refugee flows are extremely closely related, with a correlation coefficient of 0.99, so the same targeted migration patterns apply to both groups.
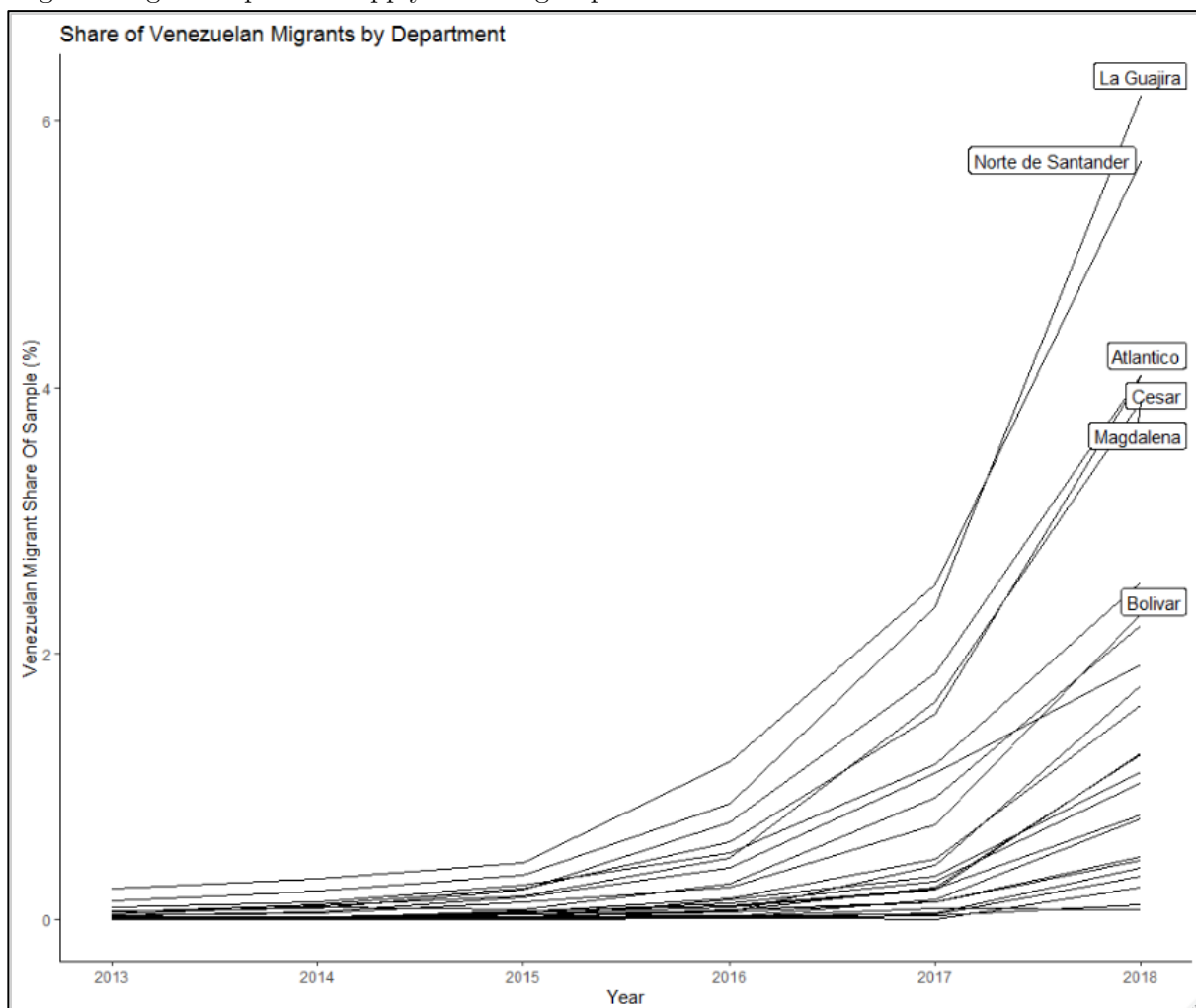
This rapid growth offers significant opportunities for analysis. While the economic context of surveyed departments is inherently different, posing an identification issue, it is

possible that differing rates of Venezuelan immigration are the most significant changes in the economic context between departments over this short time period. This dynamic motivates the use of a time-series analysis that can dynamically compare both high and low migration departments over the period of study, while a fixed-effects model can normalize for the inherent differences in departments. Inclusion of year-factor variables can also help adjust for other time effects.

Findings

Economic Characteristics of the Venezuelan Migrant and Refugee Populations

Venezuelan migrants are generally more likely than native Colombians to be in the workforce. 55% of Venezuelan migrant respondents are working and 6% are actively looking for work, compared to 47% and 3% of Colombian respondents, respectively. This trend is even more pronounced for Venezuelan refugees – 56% work and 9% are actively searching for work. Colombians are also much more likely to be students (17%) compared to either Venezuelan migrants (8%) or refugees (3%). These figures rely on self-reported activities, so the aggregate unemployment rate is significantly lower than the figures reported by the Colombian National Statistics Office (DANE), which reported a national unemployment rate of 10.5% in May 2019. The DANE methodology is less restrictive, in that it includes all individuals who did a paid activity for one hour or more in the last week – not only individuals who primarily spent their time working. This analysis will use the more restrictive definition, as the impact of migration on full-time job opportunities is the question of interest, which may be obscured by changes in one-off or gig employment.

In terms of skill level, Venezuelan refugees are comparable to native Colombians, as both groups have 25% high-skill workers. Venezuelan migrants are slightly lower, at just 18% high-skill workers. This may suggest that those who choose to migrate have different characteristics in the labour market, while native workers are more directly comparable to refugees who were forced to flee. This is further evinced by the average age of respondents – while Colombians in the sample averaged 32.6 years of age, Venezuelan refugees averaged 27.0 years and migrants 21.9 years.

Migration Analysis

Panel time series regression was modelled using the plm package in R, indexed by year and department in a fixed-effects model. The analysis found no statistically significant relationship between the share of Venezuelan migrants and the employment status, wages, or business ownership rates of native Colombians. Age, gender, skill level, and urban share were included as controls.

Relationship Between Venezuelan Migrant Share and the Economic Outlook of Colombians				
Dependent variables:	Working Share	Unemp. Share	Income	Bus. Ownership
Share of Venezuelan	-0.477	-0.038	725.505	0.506
Migrants	(-0.318)	(-0.05)	(-3,696.75)	(-0.323)
Average Age	0.083	0.1	**-10,735.510*****	-0.129
	(-0.098)	(-0.078)	**(-5,791.80)**	(-0.506)
Gender Share	**1.821*****	-0.195	1,151.30	**2.258****
(% Male)	**(-0.316)**	(-0.157)	(-11,640.54)	**(-1.017)**
Skill Share	**0.232*****	**-0.076****	3,760.61	0.044
(% High-Skill Workers)	**(-0.064)**	**(-0.038)**	(-2,809.63)	(-0.246)
Share of Urban Residents	0.055	-0.006	-1,538.84	**-0.956*****
	(-0.039)	(-0.049)	(-3,631.39)	**(-0.317)**
Observations	144	144	144	144
R^2	0.242	0.055	0.039	0.137
Adjusted R^2	0.185	-0.175	-0.195	-0.073
F Statistic	8.513***	1.333	0.941	3.651***
	(df = 5; 133)	(df = 5; 115)	(df = 5; 115)	(df = 5; 115)

Note: *p**p***p<0.01

For refugees, however, there appeared to be a significant relationship with Working Share.

Relationship Between Venezuelan Refugee Share and the Economic Outlook of Colombians				
Dependent variables:	Working Share	Unemployment	Income	Bus. Ownership
Share of Venezuelan	**-7.917****	-0.291	-35,320.20	-0.091
Refugees	**(-3.06)**	(-0.532)	(-39,200.73)	(-3.474)
Average Age	0.075	0.093	-7,895.33	0.216
	(-0.095)	(-0.079)	(-5,835.16)	(-0.517)
Gender Share	**1.914***	-0.189	1,059.66	**2.181**
(% Male)	**(-0.312)**	(-0.157)	(-11,588.11)	**(-1.027)**
Skill Share	**0.234***	**-0.078**	3,318.09	0.021
(% High-Skill Workers)	**(-0.063)**	**(-0.038)**	(-2,833.01)	(-0.251)
Share of Urban Residents	0.047	-0.005	-780.386	**-0.899***
	(-0.037)	(-0.05)	(-3,674.85)	**(-0.326)**
Observations	144	144	144	144
R^2	0.267	0.052	0.046	0.119
Adjusted R^2	0.211	-0.178	-0.187	-0.096
F Statistic	9.668***	1.270	1.102	3.094**
	(df = 5; 133)	(df = 5; 115)	(df = 5; 115)	(df = 5; 115)

Note: *p**p***p<0.01

There is an additional concern in this analysis due to nonresponse bias. Across the dataset, approximately 20% of the respondents did not answer the question regarding employment status. This creates the possibility that the changes in working share are being

driven by factors that are reflected in willingness to answer the question, and don't necessarily reflect changes in actual work status. While it is impossible to determine whether this is indeed a factor, it suggests that we should be cautious in interpreting the results.

<u>Robustness Checks</u>

If we breakout the result by department, it is clear that the effect is primarily being driven by a handful of departments that had significant changes in their Colombian working share. If the model breaks out the department fixed-effect as a factor variable, the following results emerge:

	Coefficient	Std. Error
Venezuelan Refugees	**-10.069*****	**-1.776**
Average Age	-0.005	-0.574
Gender Share	0.434	-0.45
High-skill workers	**0.414*****	**-0.113**
Urban residents	**0.741*****	**-0.142**
Atlantico	**-5.455****	**-2.118**
Bogota	**-3.798****	**-1.843**
Bolivar	-1.469	-2.109
Boyaca	-0.345	-1.741
Caldas	**-2.560***	**-1.384**
Caqueta	**6.590***	**-3.428**
Cauca	1.304	-1.512
Cesar	-1.409	-2.92
Choco	0.271	-4.222
Cordoba	2.495	-2.148
Cundinamarca	**22.057*****	**-3.764**
Huila	-0.222	-1.615
La Guajira	1.594	-4.206
Magdalena	-0.694	-2.414
Meta	2.612	-2.308
Narino	**6.161*****	**-1.441**
Norte de Santander	0.901	-2.003
Quindio	**-6.242*****	**-1.297**
Risaralda	**5.586*****	**-1.143**
Santander	**5.603*****	**-1.154**
Sucre	-0.419	-2.27
Tolima	-0.858	-0.855
Valle del Cauca	-0.468	-1.097
Observations	144	
R²	0.864	
Adjusted R²	0.824	
F Statistic	25.024*** (df = 28; 110)	

Note: *p**p***p<0.01

This result indicates a risk that the result could be spurious – driven by significant changes in a few departments, while the majority of departments show little change in either Venezuelan refugees or Colombians working. This is confirmed by robustness tests on the full model, which indicate both cross-sectional dependence and serial correlation. The Breusch-Pagan test for cross-sectional dependence and the Wooldrich test for serial correlation are both statistically significant at the highest levels. Thus it appears that the panel and time indexing in the model doesn't adequately adjust for the departmental and yearly variation within the data.

Lagged Variable & First-Difference Models

```
Oneway (individual) effect Within Model

Call:
plm(formula = l.ColoWorkingShare ~ VenezuelanMigrantShare + Age +
    MaleShare + HighSkillShare + UrbanShare, data = dpto.df,
    model = "within", index = c("YEAR", "DPTO"))

Balanced Panel: n = 24, T = 5, N = 120

Residuals:
      Min.   1st Qu.    Median   3rd Qu.      Max.
-3.490489 -0.539928 -0.014068  0.722935  3.566937

Coefficients:
                       Estimate Std. Error t-value  Pr(>|t|)
VenezuelanMigrantShare -0.54681    0.14504 -3.7700 0.0002898 ***
Age                     0.38336    0.23979  1.5987 0.1133488
MaleShare               1.30411    0.48477  2.6901 0.0084972 **
HighSkillShare          0.31873    0.11951  2.6670 0.0090571 **
UrbanShare              0.15553    0.20338  0.7647 0.4464166
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares:    175.65
Residual Sum of Squares: 127.29
R-Squared:      0.2753
Adj. R-Squared: 0.052318
F-statistic: 6.91392 on 5 and 91 DF, p-value: 1.6375e-05
```

Another possible way of understanding the temporal effects within the data is to look at the effect of Venezuelan migration on Colombian workers in the subsequent year. This lagged-variable analysis suggests that there may be a delayed impact from the migration that takes a while to appear in the department-level metrics. Unfortunately the lagged model effects were indistinguishable from the non-lagged results – migration had a statistically significant impact on native Colombians' working share, but unemployment, business ownership, and income were all unaffected. This is perhaps unsurprising, as the lagged-variable model inherently excludes the final year of the lagged variable – in this data, 2018 appears to be the

most significant year for Venezuelan migration, but its impact will not be able to be measured in the lagged model until 2019 data is available. Thus it is impossible to say at this point whether lagged impacts are significant, but hopefully this analysis can be rerun in a few years with clearer results.

Another approach is to look at the first-difference model, which attempts to predict the change in the dependent variable from period to period rather than its absolute value. This could have theoretical justification if we assume that the labour market in Colombia is robust to immigration levels in the long run, but year-by-year changes can have predictable short-term impacts before equilibrium is re-established. Once again, the results of the first-difference models were the same as for the regular time-series panel models: significant negative impacts on work share, but negligible changes in unemployment, business ownership, and income once relevant control variables are accounted for.

Effect by Skill Level

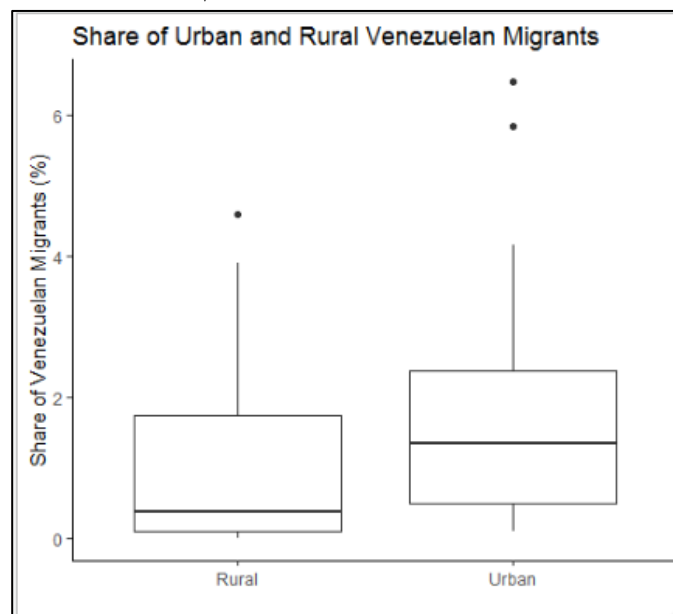|	High-Skill Workers		Low-Skill Workers	
Dependent variables:	Working Share	Unemployed	Working Share	Unemployed
Share of Venezuelan Migrants	-0.442 (-0.383)	**-0.191**** (**-0.077**)	**-1.265***** (**-0.324**)	-0.048 (-0.044)
Average Age	-0.006 (-0.348)	**0.405***** (**-0.132**)	0.11 (-0.089)	-0.013 (-0.05)
Gender Share (% Male)	**0.588***** (**-0.169**)	0.008 (-0.079)	**1.066***** (**-0.28**)	-0.133 (-0.106)
Share of Urban Residents	-0.121 (-0.093)	-0.034 (-0.115)	0.029 (-0.033)	0.021 (-0.037)
Observations	144	144	144	144
R^2	0.13	0.092	0.154	0.042
Adjusted R^2	0.072	-0.119	0.098	-0.181
F Statistic	5.016*** (df = 4; 134)	2.953** (df = 4; 116)	6.119*** (df = 4; 134)	1.267 (df = 4; 116)

Note: [*]p[**]p[***]p<0.01

The impact of immigration has distinct effects on high-skill and low-skill workers. In terms of working share, high-skill workers appear to be unaffected by the increase in Venezuelan migrants, with most of the impact appearing in the low-skill cohort. As in the general model, low-skill workers did not see any significant change in unemployment, and the high-skill workers actually saw a decrease in unemployment. This suggests that there is a possibility that increased Venezuelan migration actually serves to improve the economic prospects of high-skill workers, while negatively impacting low-skill workers. At the same time, these results should be interpreted cautiously, as the R^2 values for these models are vanishingly small – especially for unemployment – which suggests that the overall ability of this model to predict unemployment is very low.

Effect on Urban/Rural Workers

Dependent variables:	Urban Workers		Rural Workers	
	Working Share	Unemployed	Working Share	Unemployed
Share of Venezuelan	**-1.347*****	**-0.103***	-0.182	-0.049
Migrants	**(-0.314)**	**(-0.054)**	(-0.464)	(-0.042)
Average Age	-0.003	**0.183****	**-0.224***	**0.058***
	(-0.098)	**(-0.083)**	**(-0.126)**	**(-0.03)**
Gender Share	**1.803*****	-0.171	0.018	0.003
(% Male)	**(-0.301)**	(-0.147)	(-0.157)	(-0.019)
Share of High-Skill	**0.166*****	-0.042	**0.479*****	**-0.081*****
Workers	**(-0.057)**	(-0.037)	**(-0.056)**	**(-0.014)**
Observations	144	144	144	144
R^2	0.268	0.051	0.38	0.281
Adjusted R^2	0.219	-0.17	0.338	0.113
F Statistic	12.281***	1.565	20.507***	11.312***
	(df = 4; 134)	(df = 4; 116)	(df = 4; 134)	(df = 4; 116)

Note: *p**p***p<0.01

There are also divergent impacts on urban and rural workers. Urban workers see a decline in working share, and the same paradoxical decline in unemployment that is seen in the high-skill model. Rural workers, on the other hand, appear to be unaffected by the increase in migrants. Part of this divergent impact may be the difference in migration rates – according to the boxplot below, the share of Venezuelan migrants tends to be higher in urban areas. As a result, the concentration may not be high enough in rural areas to significantly impact the economy on a departmental level. There was also no significant effect on business ownership or income when broken down by urban/rural workers.

Share of Urban and Rural Venezuelan Migrants

Effect by Gender

Dependent variables:	Male		Female	
	Working Share	Unemployed	Working Share	Unemployed
Share of Venezuelan	**-0.957***	-0.034	**-0.769***	**-0.138***
Migrants	**(-0.348)**	(-0.064)	**(-0.389)**	**(-0.045)**
Average Age	0.073	**0.186***	0.119	0.099
	(-0.11)	**(-0.109)**	(-0.124)	(-0.06)
Urban Share	**0.097****	-0.018	-0.002	0.028
(% Urban)	**(-0.042)**	(-0.059)	(-0.05)	(-0.044)
Share of High-Skill	**-0.156****	**-0.110***	**0.347***	0.004
Workers	**(-0.077)**	**(-0.061)**	**(-0.07)**	(-0.028)
Observations	144	144	144	144
R^2	0.085	0.032	0.203	0.08
Adjusted R^2	0.024	-0.194	0.149	-0.134
F Statistic	3.127**	0.947	8.529***	2.513**
	(df = 4; 134)	(df = 4; 116)	(df = 4; 134)	(df = 4; 116)

Note: *p**p***p$<$0.01

Gender seemed to mirror the overall dynamic in the general model, with a measurable decline in working share but no significant changes to other economic variables. There was an apparent decrease in female unemployment, but again the vanishingly small R^2 value suggests that the overall ability of the model to make predictions is extremely limited. Thus it appears that the effect is not particularly pronounced based on gender.

Conclusion

Early indications suggest that there may be pressures on the labour market for native Colombians as a result of the influx of Venezuelan migration. This effect is especially pronounced for low-skill and urban workers, suggesting that Venezuelan refugees and migrants compete with native Colombians for low-skill jobs, causing Colombians to stop working. It appears like these workers drop out of the workforce entirely, as there is no corresponding uptick in job-searching. The overall predictive power of these models is relatively low, however, precluding their use as tools of policy design.

Limitations

In addition, the analysis faced several constraints that limited the impact of its results. The wave of migration began to increase substantially in 2016, limiting both the overall size of the relevant dataset and the ability to perform time-indexed analysis like first-difference and lagged-variable models. The dataset was also limited in its geographic index – statistics had to be aggregated at the department level as it was the most geographically specific index that was available. This limited the ability to make meaningful claims, as department-level statistics incorporate numerous endogenous factors that are not reflected in the analysis. Future research should seek to identify sub-Departmental opportunities to further analyse this phenomenon. As more years of GEIH data become available, there should also be opportunities to examine the long-run impact of the Venezuelan crisis.