Essays on Immigration, Labour Markets, and Political Behaviour

Diala Al Masri

St. John's College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy in International Development

September 2022

Essays on Immigration, Labour Markets, and Political Behaviour

Diala Al Masri

St. John's College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy in International Development

September 2022

Abstract

This thesis is composed of three self-contained empirical chapters that tackle the labour market implications from COVID-19, and the political and trade implications from recent immigration, in the Brazilian context. The first chapter presents a short and medium term analysis of the impact of the COVID-19 pandemic on the Brazilian labour markets. I find that informal workers, females, youth, and workers in small firms are the ones most adversely impacted in the short term. In the medium term, I show that there seems to be a lingering negative effect on women's employment and wages, while informal workers seem to recover much faster than their formal peers. I further study the role of the government response in mitigating the shock of the pandemic, and the distributional impacts under different counterfactual targeting scenarios of the generous but temporary cash transfer program, Auxílio Emergencial. I find that the Auxílio Emergencial transfers more than offset the labour income losses for the bottom four deciles, such that inequality and headcount poverty ratio decreased relative to the pre-COVID baselines. The second chapter explores the impact of international immigration flows, exploiting the spatial variation across Brazilian municipalities from 2010 to 2017. I estimate the correlation between immigration and trade using a rich set of fixed effects, motivated by an augmented gravity model. I then provide a causal estimate of the impact using a Bartik shift-share instrumental variable approach. I find evidence of a pro-trade creation effect, particularly on differentiated imports, at the intensive and extensive product levels. I document that this impact differs depending on certain characteristics of the immigrants' country of origin. The third paper analyses the impacts of the Haitian immigration on the voting outcomes and electoral participation of citizens in Brazil's 2018 Presidential elections. I exploit this quasi-natural experiment, and I use a variation on the shift-share instrument to arrive to a causal impact. I find that the presence of Haitian immigrants led to an overall average decrease in political participation across different education groups and an increase in

vote shares towards right-wing parties. I analyse potential channels through which immigrants influence voting behaviour. While I rule out the mechanisms of natives' fear over public resources, natives' flight (voting with their feet), and changes in candidate composition, I present evidence that suggests that the change in voter behaviour is happening through cultural and economic mechanisms.

Supervisors This thesis was supervised by Professor Douglas Gollin and Professor Gabriel Ulyssea.

Total words This thesis contains approximately 95,122 words. This word count has been calculated by counting the number of words on a representative page (478 words, page 7) and multiplying this with the number of pages, excluding the cover, table of contents, bibliography and appendices.

Statement of authorship Chapter one is presented as a sole authored project. I note that it builds, expands and innovates on an original joint project written during and after my internship at the International Monetary Fund. I gratefully acknowledge my collaborators in the original project: Valentina Flamini, Fredrick Toscani, and Carlo Pizzanelli. Chapter two is presented in my thesis as a sole authored project, and is an expanded and completely revised version of a project that has a co-authored antecedent submitted as a separate project to the Inter-American Development Bank. I gratefully acknowledge my collaborator in the consultancy project, Professor Carlos Vargas Silva. Chapter three is sole authored and has no co-authored antecedents. I have received approval from the relevant collaborators and the Graduate Studies Committee to use the chapters under the declared sole authorship.

Acknowledgment of funding I gratefully acknowledge the generous funding from the Rhodes Trust.

To my mother, Yusr.

Acknowledgements

I dedicate this thesis to my mother, Yusr. I am fulfilling her dream by pursuing this Ph.D., and seeing it to completion. I will remain all my life, in a huge debt of gratitude, for her belief in me, selfless sacrifices, courage and resilience, and her unwavering love for life, singing, and writing. All what is good in me, I learned it from you, mom.

I am grateful to my supervisors, Professor Douglas Gollin and Professor Gabriel Ulyssea. I will continue to grow into the economist you inspire me to be. Doug, your steady presence, patience, and holistic support has been indispensable for my growth as a professional and as a person. I will pay it forward. Gabriel, I learned a tremendous amount from your supervision throughout my MSc., and Ph.D., and your insights and expertise were crucial to the completion of this work. I also want to thank Professors Christopher Adam, Peter Montiel, and Peter Pedroni for being so generous with their teachings.

I am beyond blessed to have the support of my family. Without them, it would have been impossible for me to be where I am now. Rami, Jado, and my father, Issam: Thank you for holding space for my Ph.D., despite the many more pressing issues in the last few years, imposed on you by living in Lebanon. I am ever so grateful for your unconditional love and understanding, and for being my closest friends. Toufic: I can't wait for the new chapters together. Thank you for your tireless encouragement, understanding, and steadiness in pushing me forward. I am better everyday because of your presence.

To my friends in Oxford, Lebanon, and beyond, I am thankful for your unwavering support and unconditional presence. You made a home for me. Abeer, thank you for always making space for all what I have to share. You make Lebanon feel closer. Maureen, thank you for all the care, for bringing me into your wonderful family, and for helping me find my way back to my sense of self and curiosity. Margaret and Magnus, thank you for being there throughout it all, for being a home for me, and for being a source of support and comfort. Gaby and Kimi, I couldn't have imagined anyone else to share this Ph.D. journey with. You made everyday brighter, for that, for your sisterhood, and for much more, thank you. My St. John's dolphins Razia and Barbara, from junior deans to friends to family, I am so grateful for you. Ashley, Farah, Tatjana, Lisa: I will never forget all the times you stood by my side and all our memories together.

Contents

Acknowledgements **v**

Introduction **1**

I The Impact of COVID-19 on Labour Markets and the Distributional Impacts of Auxílio Emergencial **13**
1 Literature Review 20
2 Data 23
3 Analysis of the Impact of the Pandemic on Labour Markets 24
4 The Impact of the Government's Response to the Pandemic 60
5 Conclusion 87

Appendices of Chapter 1 **91**
I.A Supplementary Material: Ex-Ante Accounting Analysis 92
I.B Supplementary Material: Employment Statistics During the Pandemic . 96
I.C Supplementary Material: Auxílio Emergencial 99
I.D Supplementary Material: Containment Measures 114

II Immigration and Trade in a Middle-Income Country **124**
1 Introduction 126
2 Migration and Trade 130
3 Data 136
4 Channels of Impact and Empirical Evidence 140
5 Estimating a Causal Impact 152
6 Limitations of the Analysis 172
7 Conclusion and Policy Implications 174

Appendices of Chapter 2 **178**
II.A Supplementary Material: Validity Tests of the Shift Share IV 179
II.B Supplementary Material: Descriptive Statistics 187
II.C Supplementary Material: Robustness Checks and Different Specifications for Results 190

III Haitian Immigration and Citizens' Voting Behaviour **227**
1 Introduction 229
2 Literature Review and Channels of Impact 236
3 Background 244
4 Data 257
5 Empirical Strategy 259
6 Results 271
7 Conclusion 300

Appendices of Chapter 3 **305**
III.A Supplementary Material: Ideology Categorisations 306
III.B Supplementary Material: IV Shift Share Robustness Checks 308
III.C Supplementary Material: Robustness Checks related to Data Treatment 326
III.D Supplementary Material: Descriptive Analysis 333
III.E Supplementary Material: Additional Results and Robustness Checks . . 341
III.F Supplementary Material: Case Study of Venezuelan Immigration 363

Conclusion **373**

References **376**

List of Figures

I.1 Epidemiological Situation in Brazil 26
I.2 General Overview 27
I.3 Key Labour Market Indicators 27
I.4 Teleworkability and Contact-Intensity Index by Sector 29
I.5 Overall Ex-ante Employment-at-Risk across Stages and by Sector . . 40
I.6 Employment and Wage-at-Risk in the Initial Lockdown Stage 41
I.7 Decomposition of Shock Type by Income Deciles and Formality Status 41
I.8 Ex-ante Employment-at-Risk 42
I.9 Ex-ante vs Ex-post Employment Changes 44
I.10 Ex-ante vs Ex-post Employment Changes 44
I.11 Ex-ante vs Ex-post Employment Changes by Firm Size 45
I.12 % Change in Employment, hours worked and monthly income 47
I.13 Decomposition by Employment Categories 48
I.14 Formal and Informal Work 48
I.15 Decomposition by Age Categories 49
I.16 labour Market Impacts by Gender 51
I.17 % Change in Employment, hours worked and monthly income 51
I.18 Pre- and Post-COVID Hours Worked and Wages Across Sectors . . . 53
I.19 Pre- and Post-COVID Hours Worked and Wages Across Occupations 53
I.20 Employment, Weekly Hours, and Monthly Wage for Rural and Urban Regions 54
I.21 Stringency and Economic Support Indices 62
I.22 Percent Change in Employment by Sector in Response to Government Measures 70
I.23 Distribution Analysis based on post-COVID Labour Income 74
I.24 Share of Total Households and Share of Total EA Paid per Decile . . 78
I.25 Income Distribution with EA by Income Decile 78
I.26 Poverty Headcount Ratio (%) and Inequality 84
I.27 Female Poverty Headcount ratio Counterfactual Analysis for the Administration of EA 85
I.28 Poverty Counterfactual Analysis for the Administration of EA 87
I.29 Employment Decomposition in May 96
I.30 Remote Employment by Firm Size in May 97

I.31 Distribution of Workers across Sectors with Margins of Heterogeneity 97
I.32 Percentage of Households by Aid Value 103
I.33 Visual Discontinuity in the Running Variable 108
I.34 Bunching Around Threshold 111
I.35 Placebo Thresholds 112
I.36 Percent Change in Formal and Informal Employment by Sector in Response to Government Measures 118

II.1 Immigration by Year 131
II.2 Distribution of All Migrants in Brazil 131
II.3 Trade in Brazil 135
II.4 Descriptive Graphs for disaggregated Trade Variables by Sector Elasticity of Substitution σ and Rauch Classification 136
II.5 Impact of Migration on Trade: Costs and Preferences Channels 144
II.6 Heterogeneous Impact on Trade by Country of Origin Characteristics 171

III.1 The Migratory Flow and Entry Points of Haitians 245
III.2 The Distribution of Haitian Immigrants 248
III.3 Vote Shares and Ideology of Presidential Parties Across Time 253
III.4 Votes for Left Wing Parties 254
III.5 Ideological Positions of the Top Four Presidential Candidates 256
III.6 Proximity to Haitian Creole $\lambda_{f,creole}$ by Historical Immigrant Spoken Languages (before 2000) 265
III.7 Effects of Haitian Immigration on Vote Shares by Mobilising or Switching Effects 280
III.8 Results on Individual Turnout 292
III.9 Correlation of Opinions on the Most Important Problem with Haitian Presence and Voter Turnout 293
III.10 Results from Multinomial Logistic Regression 295
III.11 Distribution of the Haitians and the calculated IV across municipalities 308
III.12 Distribution of the Haitians and the calculated IV within Mesoregions 309
III.13 Testing for Pre-trend Correlations 317
III.14 Discussion on the IV Share Component 322
III.15 Robustness Checks: Power of IV under Different F.E and Levels of Clustering 323
III.16 Distribution of Traditional IV 324
III.17 Distribution of Haitian Immigrants 333
III.18 Votes for Haddad, Alckmin, and Bolsonaro in 2018 337
III.19 Proximity to Haitian Creole $\lambda_{f,creole}$ by Country 341
III.20 Distribution of Venezuelan Immigrants and Main Municipalities of Entry 365
III.21 Synthetic Control Fit 368
III.22 In-Place Treatment Effects 369

List of Tables

I.1 Labour Income and Hours Worked by Income Decile 29
I.2 Labour Market characteristics, Demand Shock and Probability of Lockdown by Vulnerability to Lockdown 36
I.3 Worker Characteristics and Job Attributes by Vulnerability to Lockdown 37
I.4 Employment-at-Risk by Firm size (in millions) 43
I.5 Heterogeneity of Impact by worker and industry characteristics 57
I.6 Continued: Heterogeneity of Impact by worker and industry characteristics 58
I.7 Impact on labour Income and Hours Worked 66
I.8 Heterogeneity of the Impact of Government Response on Hours Worked and Labour Income 68
I.9 Timeline of the three Auxílio Emergencial (EA) Phases 73
I.10 Baseline Characteristics of Individuals Receiving Auxílio Emergencial 76
I.11 Impact on labour Income and Hours Worked 82
I.12 The calibration measures of $Lock_s$ and $Lock_{s_{selective}}$ 92
I.13 S_{it} Values Across the Shutdown Stage and the Two Opening Stages . 93
I.14 $Direct_{Loss_s}$ (Direct Demand Shock) 93
I.15 Distribution of Workers by Sex and Education and Occupation Teleworkability and Contact Intensity 94
I.16 Characteristics and Job Attributes for Formal and Informal Workers . 95
I.17 Change in Income and Hours Worked 98
I.18 Eligibility and Benefits of Auxílio Emergencial 100
I.19 Distribution of Auxílio Emergencial across Federation Units as a % of Total Population 101
I.20 Distribution of Auxílio Emergencial across Federation Units as a % of UF Population 102
I.21 Aid Received by Top 2 Deciles Households 103
I.22 Balance Tests for PNAD COVID 105
I.23 First Stage 109
I.24 Balance Test on Observable Characteristics 110
I.25 Manipulation Test 111
I.26 Results on Labour Market Outcomes in November 113
I.27 Balance Tests for PNAD Continua Matched Sample 117

I.28 Average of Main Aggregate Indices across Time by Federation Units . 120
I.29 Average of the Subcomponents of the Containment Measure across Time by Federation Units 121
I.30 Average of the Subcomponents of the Economic Support Measure across Time by Federation Units 122
I.31 Average of the Subcomponents of the Stringency Measure across Time by Federation Units 123

II.1 Average Trade Across Municipalities 139
II.2 Locations of All Migrants Across Brazil by Year 139
II.3 Effect of Immigration on Trade (in USD) based on Rauch Classification 149
II.4 Effect of Immigration on Trade (in USD) based on σ Elasticity of Substitution 150
II.5 **Extensive Margin:** Effect of Immigration on Trade (in USD) based on Rauch Classification 150
II.6 **Extensive Margin:** Effect of Immigration on Trade (in USD) based on σ Elasticity of Substitution 151
II.7 First Stage Estimates 158
II.8 Total Exports and Imports 164
II.9 Differentiated and Homogeneous Exports and Imports based on Rauch Classification 164
II.10 **Extensive Margin:** Differentiated and Homogeneous Exports and Imports based on Rauch Classification 165
II.11 **Intensive Margin:** Differentiated and Homogeneous Exports and Imports based on Rauch Classification 165
II.12 Based on σ Elasticity of Substitution 166
II.13 **Extensive Margin:** Based on σ Elasticity of Substitution 166
II.14 **Intensive Margin:** Based on σ Elasticity of Substitution 167
II.15 Evaluation of Underlying Shares 181
II.16 Correlations with the Highest Rotemberg Language Shares and Potential Covariates 182
II.17 Correlations in Initial Period Covariates and Migration Decisions in the Subsequent Decade 182
II.18 Pre-trends correlations 183
II.19 Correlations Across Cohorts for All Language groups 185
II.20 Correlations Across Cohorts for Top 14 Language Groups 185
II.21 First Stage Estimates for All Migrants Based on Country of Origin . 186
II.22 Immigrants by Country of Origin: Total and Growth Rates (%) . . . 187
II.23 Migrant Characteristics by Nationality Across All Years 188
II.24 Description of Dependent and Independent Variables 189
II.25 Exports and Imports disaggregated by σ Interaction 193
II.26 Effect of Immigration on Trade based on Rauch Classification (USD) . 194
II.27 **Extensive Margin**: Effect of Immigration on Trade based on Rauch Classification (USD) 195

II.28 Effect of Immigration on Trade based on σ Elasticity of Substitution (USD) . . . 196
II.29 **Extensive Margin:** Effect of Immigration on Trade based on σ Elasticity of Substitution (USD) . . . 197
II.30 Effect of Immigration on Trade based on Rauch Classification . . . 198
II.31 **Extensive Margin:** Effect of Immigration on Trade on the Extensive margin (in USD) based on Rauch Classification . . . 199
II.32 Effect of Immigration on Trade (in USD) based on σ Elasticity of Substitution . . . 200
II.33 **Extensive Margin:** Effect of Immigration on Trade (in USD) based on σ Elasticity of Substitution . . . 201
II.34 Based on Rauch Classification . . . 202
II.35 Based on σ Elasticity of Substitution . . . 202
II.36 Based on Rauch Classification on the Extensive margin . . . 203
II.37 Based on σ Elasticity of Substitution on the Extensive margin . . . 203
II.38 Total Exports and Imports . . . 204
II.39 Differentiated and Homogeneous Exports based on Rauch Classification 204
II.40 Differentiated and Homogeneous Imports based on Rauch Classification . . . 205
II.41 Based on σ Elasticity of Substitution . . . 205
II.42 Total Exports and Imports . . . 206
II.43 Differentiated and Homogeneous Exports based on Rauch Classification 206
II.44 Differentiated and Homogeneous Imports based on Rauch Classification . . . 207
II.45 Based on σ Elasticity of Substitution . . . 207
II.46 **Excluding Outlier Municipalities:** Total Exports and Imports . . 209
II.47 **Excluding Outlier Municipalities:** Differentiated, Homogeneous, and Reference Exports based on Rauch Classification . . . 210
II.48 **Excluding Outlier Municipalities:** Differentiated, Homogeneous, and Reference Imports based on Rauch Classification . . . 210
II.49 **Excluding Outlier Municipalities:** Based on σ Elasticity of Substitution . . . 211
II.50 **Excluding non-Trading Municipalities:** Total Exports and Imports . . . 211
II.51 **Excluding non-Trading Municipalities:** Differentiated, Homogeneous, and Reference Exports based on Rauch Classification . . . 212
II.52 **Excluding non-Trading Municipalities:** Differentiated, Homogeneous, and Reference Imports based on Rauch Classification . . . 212
II.53 **Excluding non-Trading Municipalities:** Based on σ Elasticity of Substitution . . . 213
II.54 **Excluding Border Municipalities:** Total Exports and Imports . . 213
II.55 **Excluding Border Municipalities:** Differentiated, Homogeneous, and Reference Exports based on Rauch Classification . . . 214

II.56 **Excluding Border Municipalities:** Differentiated, Homogeneous, and Reference Imports based on Rauch Classification 214
II.57 **Excluding Border Municipalities:** Based on σ Elasticity of Substitution 215
II.58 **Controls in year 2000 interacted with a time trend:** Total Exports and Imports 215
II.59 **Controls in year 2000 interacted with a time trend:** Differentiated, Homogeneous, and Reference Exports based on Rauch Classification 216
II.60 **Controls in year 2000 interacted with a time trend:** Differentiated, Homogeneous, and Reference Imports based on Rauch Classification 216
II.61 **Controls in year 2000 interacted with a time trend:** Based on σ Elasticity of Substitution 217
II.62 **Less than 15 immigrants historically:** Total Exports and Imports 218
II.63 **Less than 15 immigrants historically:** Differentiated and Homogeneous Exports based on Rauch Classification 219
II.64 **Less than 15 immigrants historically:** Differentiated and Homogeneous Imports based on Rauch Classification 219
II.65 **Less than 15 immigrants historically:** Differentiated and Homogeneous Exports and Imports based on Rauch Classification on the Extensive Margin 220
II.66 **Less than 15 immigrants historically:** Based on σ Elasticity of Substitution 220
II.67 **Less than 15 immigrants historically:** Based on σ Elasticity of Substitution on the Extensive margin 221
II.68 **Mercosur:** Total Exports and Imports 221
II.69 **Mercosur:** Differentiated and Homogeneous Exports and Imports based on Rauch Classification 222
II.70 **Mercosur:** Based on σ Elasticity of Substitution 223
II.71 **Mercosur:** Differentiated and Homogeneous Exports and Imports based on Rauch Classification at the Extensive Margin 223
II.72 **Mercosur:** Based on σ Elasticity of Substitution at the Extensive Margin 224
II.73 **OECD:** Differentiated and Homogeneous Exports and Imports based on Rauch Classification 224
II.74 **OECD:** Based on σ Elasticity of Substitution 225
II.75 **OECD:** Differentiated and Homogeneous Exports and Imports based on Rauch Classification at the Extensive Margin 225
II.76 **OECD:** Based on σ Elasticity of Substitution at the Extensive Margin 226

III.1 Migrant Characteristics by Nationality Across All Years 250
III.2 The Political Federal System in Brazil 252
III.3 First Stage Results 269

III.4 First Stage Results: Robustness Checks I 270
III.5 First Stage Results: Robustness Checks II 270
III.6 Turnout Ratio Results 273
III.7 Turnout Ratio Results by Education Level and Gender 274
III.8 Disaggregated Party Classifications: IV 2SLS Estimates for First Round Elections 277
III.9 Vote Shares to Presidential Candidates: IV 2SLS Estimates for First Round Elections 277
III.10 Channels of Impact: Welfare-state and Native Flight Outcomes 288
III.11 Channels of Impact: Economic Outcomes 290
III.12 Descriptive Statistics on Presidential Voting Choices by Individual's Preferences on Immigration Policies 291
III.13 Brazilian Party Codes and Ideology Classifications 307
III.14 Estimates on centrist vote shares 311
III.15 Rotemberg Weights and Contribution of Language Shares 313
III.16 Correlation of Shares with Municipality Characteristics in 2010 315
III.17 Language Share impact on Pre-trends 318
III.18 Pre-Trends Correlations on Predictor Variables 319
III.19 Pre-Trends Correlations on Outcome Variables 319
III.20 IV Predicting Non-Haitian Immigrants 320
III.21 IV Leave one-out strategy at the Microregion 322
III.22 IV First Stage with Traditional IV 324
III.23 IV First Stage with the Language Proximity Index using All pre-2000 Immigrants 325
III.24 First Stage Results: Haitian Migration Robustness 328
III.25 First Stage Results: Haitian Migration Without Asylum Seekers . . . 330
III.26 Balance test on SISMIGRA dataset: Immigrants with Missing versus Not Missing Location 331
III.27 Balance test on STI MAR dataset: Immigrants with Missing versus Not Missing Location 332
III.28 Haitians' Municipality Characteristics per Year 334
III.29 Number of Flows and Visa Types of Haitian Immigrants 334
III.30 Electoral Outcomes: Balance tests between Haitian receiving and non-receiving municipalities 335
III.31 Control Variables: Balance tests between Haitian receiving and non-receiving municipalities 336
III.32 Language Assigned by Country 341
III.33 Holding Turnout Ratio Fixed: Results across bands of turnout ratio . 342
III.34 Impact of Haitian Immigration on Vote Shares to Presidential Candidates (by Number of Voters)- 2SLS estimates 343
III.37 Turnout Ratio: General 343
III.35 Impact of Haitian Immigration on Blank Votes - 2SLS estimates . . . 344
III.36 Impact of Haitian Immigration on Null Votes - 2SLS estimates 345
III.39 Turnout Ratio: By Education and Gender (First Round) 346

III.38 Turnout Ratio: Compulsory versus Voluntary 347
III.40 Impact of Haitian Immigration on Turnout Ratio: By Education and Gender (Second Round) 348
III.41 Heterogeneous Impacts on Presidential Votes 349
III.42 Heterogeneous Impacts on Voter Turnout 350
III.43 Impact of Haitian Immigration on Heterogeneous Impacts on Presidential Votes based on Visa Status 351
III.44 Heterogeneous Impacts of Haitian Immigration on Voter Turnout by Different Visa Holders 352
III.45 Impact of Haitian Immigration on Turnout Ratio: Heterogeneity by Education and Age 353
III.46 Impact of Haitian Immigration on Turnout Ratio for Females: By Education and Age for Females 354
III.47 Impact of Haitian Immigration on Turnout Ratio for Males: By Education and Age for Males 355
III.48 Turnout Ratio: Heterogeneity by Sex and Age 356
III.49 Sample of municipalities that received Haitian immigrants and those that received zero immigrants 357
III.50 Sample with only Haitian receiving municipalities and non-migrant receiving municipalities 358
III.51 Channels of Impact: Natives' Internal Migration 360
III.52 Descriptive Statistics on Presidential Turnout by Individual's Preferences on Immigration Policies 361
III.53 Impact on Candidate Characteristics 362
III.54 Synthetic Pacaraima Root Mean Squared Prediction Error 366
III.55 Weights and Predictor Means 367
III.56 Synthetic Control: Distribution Effects 370
III.57 Synthetic Control: Comparing Results of Haitian versus Venezuelan Entry Municipalities 372

Introduction

There are not many topics that have taken central stage in policy discussions and engaged public discourse in recent years as much as the implications arising from the COVID-19 pandemic and those as a result of immigration. Both of these themes have wide economic and political implications both in the short and long run. The COVID-19 pandemic has created an economic impact of unprecedented scale and, through differential susceptibility to this crisis, revealed and created large inequalities among individuals. Responses to the pandemic have depended to a large extent on the fiscal space available for a country and on the existing social support policies in place – and on political support for expansive government programs. This in turn relates to perceptions of who benefits from the largesse of government spending and social programs. In recent years, large flows of migrants across borders have raised political sensitivities, and populist politicians in many countries have sought to use immigration as a wedge issue to attack existing governments – and sometimes to undermine the legitimacy of re-distributive policies. Waves of forced immigration from Syria, Haiti, Venezuela, Afghanistan, and Ukraine, among others, have posed a challenge to existing migration policies in destination countries, and immigration has become a major political campaigning point.

Inspired by the timeliness and importance of these two themes – the pandemic and international migration flows – my thesis is composed of three empirical chapters that touch closely on these issues and their interconnections. The first paper analyses the labour market impacts of the pandemic, while the second and third papers analyse the trade and political impacts of recent immigration on economies that are destinations for recent immigration flows. All three chapters are self-contained pieces of research and are quantitative in nature.

Not only are the two themes studied in this thesis related through their economic and political consequences, but they are also studied within the Brazilian context. Being the most populous Latin American country with the largest economy (in total

GDP), Brazil represents an ideal context in which to analyse the impacts of the COVID-19 pandemic and immigration on labour markets and political economy. Following in the wake of a major recession, from which the country had only recently begun to recover, the COVID-19 pandemic brought a large share of the Brazilian economy to a halt. Brazil ranked second in the world in the number of COVID-19 cases in mid-2020 and accounted for a quarter of the global deaths (Hallal, 2021). At that time, Brazil allocated R$ 42 billion (USD 7.8 billion) specifically to address the ramifications from the COVID-19 crisis. The central government introduced the generous cash transfer program Auxílio Emergencial to mitigate the adverse effects of the pandemic on inequality and poverty. Prior to the onset of the pandemic, several factors helped attract immigrants to Brazil, including Brazil's economic prosperity in the years after 2003, its strategic international presence (for example, the humanitarian and peace keeping missions in Haiti), and the hosting of major sports events, such as the FIFA World Cup in 2014 and the Olympics and Paralympics in 2016.[1] Immigrants also arrived from distant countries and from countries that were not part of the historically traditional profile. Immigrants were absorbed by job offers initially, but many got their offers revoked following the onset of the recession in 2014. Most notable among the changes in the makeup of immigrants to Brazil was the large-scale influx of immigrants from western hemisphere countries that suffered from humanitarian, political or economic crises, such as Haiti and Venezuela. As a result, questions surrounding the impacts of immigrants on the economy and on political behaviour came to the forefront of attention of policy makers as well as of the academic, political, and electoral debates.

In the first chapter of my thesis, I conduct an extensive analysis of the impact of the pandemic, taking into consideration the specific sectoral composition of the Brazilian economy and studying the effects of the government's response on the recovery of employment, labour income and hours worked. I also study the distributional impacts of the generous social protection system implemented to combat the pandemic. My analysis has two major themes. In the first theme, I analyse the impact on labour market outcomes within a conceptual framework, following closely, with a few modifications, the work of Alfaro et al. (2020). In this framework, I take into account

[1]Brazil's foreign policy in recent decades contributed to its international presence. Part of this foreign policy is the *Política Externa Migratória Brasileira (PEXM)* that influenced immigration into Brazil from Latin America and West Africa. As part of PEXM, there has been certain strategic policies that aimed to promote the country abroad such as 1- participation in humanitarian missions in Haiti and some African countries, scholarships for individuals from Latin America and Africa, and the Mais Médicos (More Doctors) Program (Uebel, 2020; Moreira, 2015).

the distribution of workers across firm sizes, occupations, sectors, and formality status. I analyse the labour impact of the pandemic across three different phases of the evolution of the pandemic and then compare the predictions from the framework with actual labour market outcomes in the short term (i.e., the second quarter of 2020). Then, I empirically investigate the heterogeneous impact of the pandemic depending on worker, sector and occupation characteristics in the early stages of the pandemic as well as in the medium term to shed light on recovery and scarring effects.

The second part of my study analyses the role of government containment measures and the distributional impact of Auxílio Emergencial. I study the complex government response in two steps. First, using the Oxford COVID-19 Government Response Tracker (OxCGRT), I analyse the impact that containment measures and the overall level of economic support had in assisting the recovery of labour markets and sectoral employment in the subsequent quarter, focusing on the second and third quarters of 2021. Second, I study the effects that Auxílio Emergencial had on short-term poverty and inequality, based on monthly survey data from PNAD COVID.[2] I simulate how the distributional effects would have changed under different and costly targeting counterfactuals.

My findings from the first chapter are as follows. I document that the recession induced by the pandemic presented new margins of adjustment compared to the earlier 2014-2016 recession, and that the impact was highly unequal across individuals. Unlike its previous cushioning role in the earlier recession, the informal sector shrank in the short term as a result of the pandemic. Moreover, there was a much larger increase in discouraged workers and other people outside the labour force in comparison with the earlier recession. The unequal impact of the pandemic emerges across multiple margins of heterogeneity. Employment fell by over 18-34 percent year-on-year in vulnerable sectors such as construction, household services, and hospitality. The reduction in employment disproportionately affected informal workers, women, workers in small firms, and young cohorts of workers. Older cohorts of workers retained their employment but were more likely to be temporarily removed from their jobs, with high reductions in hours worked and monthly wages. Households in the bottom deciles of the income distribution were most affected, with extreme poverty and the Gini coefficient based

[2]My analysis relies on two household level surveys: PNAD-Continua and PNAD COVID, which special edition of the PNAD-Continua household survey. Both surveys are representative at the 27 federation unit levels and at the national level, and are compiled by the Instituto Brasileiro de Geografia e Estatistica (IBGE). PNAD stands for *Pesquisa Nacional por Amostra de Domicílios - PNAD*or National Household Sample Survey.

on labour income increasing sharply when measured at the height of the pandemic in May 2020. In the medium term, in the third quarter of 2021, women, youth and workers in small firms suffered persisting adverse effects on their labour income and hours worked. In contrast, informal workers were more negatively impacted than their formal counterparts, but their wages and weekly hours worked recovered more quickly than those of their formal peers. I also document a recovery in employment for the younger cohort of workers (between the ages 14-23 years old) after an initial significant reduction in employment at the onset of the crisis.

Further, based on the analysis of the impact of governmental measures, I find that contemporaneous lockdown measures led to a reduction in labour income by 0.23% and reductions in sectoral employment concentrated mainly in education and household workers, while lagged measures of lockdown stringency were positively correlated with employment, working hours and sectoral employment in the subsequent 3rd quarter of 2021. With respect to Auxílio Emergencial (EA), I find that EA more than compensated the negative impact of the pandemic on poverty and income inequality, at least temporarily. In the absence of EA, inequality as measured by the Gini index would have increased from 0.53 pre-COVID to 0.58 post-COVID, not least because Brazil's large informal sector was deeply affected by the pandemic. Likewise, the poverty headcount ratio would have increased sharply from about 22% to 30%, corresponding to approximately 16.8 million individuals.[3] Instead, with the EA implementation, the poverty headcount ratio fell to 4.4 percent and the Gini coefficient to 0.51—both lower than their pre-COVID levels. Finally, in a counterfactual analysis based on survey data for October, 2021, I simulate the redistribution power of the EA under different targeting and benefit options. I find that with better targeting, just two thirds of the observed transfer would have led to the same extreme poverty rate as the one observed in October, even after accounting for the administrative cost of better targeting. This shows the room for a more cost-effective design should a renewed emergency cash transfer be implemented.

The second and third chapters focus on studying the impacts that immigrants have on their host communities. Brazil was the destination for a diverse set of immigrants that differ in terms of their linguistic and cultural proximity to Brazilians. Brazil's municipalities experienced heterogeneous exposure to immigrant flow episodes, and hence a differential concentration of the ratio of immigrants to population. Moreover, there

[3]This analysis is based on the upper middle income poverty line of USD 5.5 per day. Based on the extreme poverty line of USD 1.9 per day, the headcount poverty ratio would have shot up from 4.7% to 13.9%.

could be heterogeneity in the response of firms and the native population depending on characteristics of the latter. Finally, recent waves of forced migration present themselves as potential quasi-natural experiments that lead to an exogenous increase in the migrant population of Brazil that can help causally identify the impact of interest. The data on immigration relies on databases released by the Observatory of International Migration (OBMigra).[4] This data is relatively new and updated regularly and I am one of the early people to start analysing this data (Galvao and Uebel (2019); Tonhati and De Macedo, (2021); frequent reports by OBMigra), especially to answer economic questions. The dataset makes it possible to establish the profile of each immigrant, including sex, country of birth and municipality of of residence, year of entry and year of registration. I complement this database with another that reports data on asylum requests reported by the International Traffic System Alerts and Restrictions Module (STI-MAR), under the responsibility of the Federal Police.

In my second chapter, I explore the impact of international immigration on trade, exploiting the spatial variation across Brazilian municipalities from 2010 to 2017. I first approach this research question by discussing the potential channels through which immigration impacts trade, based on Tai (2009); I then empirically link the theoretical predictions to the data. I estimate the impact on traded goods with varying degrees of product differentiation through interacting migrant stock in each municipality with the respective goods' elasticities of substitution. I hypothesise that immigrants can potentially have the highest impact on preferences for highly differentiated goods, and the highest reduction in fixed costs for the less differentiated ones, due to the more straightforward transaction costs in the latter.

Then, I proceed to provide empirical evidence on the relationship between immigration and trade. I begin with estimating the relationship, at the country, year and municipality levels, motivated by the augmented gravity equation based on Tai (2009). The dependent variable is exports or imports from and to Brazilian municipalities. I regress these trade outcomes on the share of immigrants per population. I estimate this empirical analysis using Ordinary Least Squares (OLS) and Psuedo Poisson Maximum Likelihood (PPML) with a rich and demanding set of fixed effects that help control

[4]The Observatory of International Migration (Observatório das migrações internacionais , OBMigra) is an institute for the purpose of studying international migratory patterns in Brazil, in cooperation with the Ministry of Labor and Employment (Ministério do Trabalho e Emprego , MTE), the National Council for Immigration (Conselho Nacional de Imigração , CNIg), and the Department of Latin-American Studies (Departamento de Estudos Latino-Americanos , ELA) of the Social Sciences Institute (Instituto de Ciências Sociais , ICS) at the Universidade de Brasília (UnB) (UnB n.d.)

for important variables that could affect the relationship, such as common language or borders, distance, trade agreements, colonial relationship, among others (Anderson and van Wincoop, 2003; Feenstra, 2003; Helpman et al., 2008; Head and Mayer, 2014).

Despite the rich set of fixed effects employed in the estimation, however, this approach remains subject to endogeneity concerns because immigrants do not locate randomly within a host country. They are likely to choose the locations that are favourable to them, economically, socially, politically, and otherwise. Also, a concern of reverse causality persists, as trade itself could be a catalyst for more immigration. Therefore, other studies on the impacts of immigration, including those studying the impacts on trade (e.g., Peri and Requena-Silvente 2010 and Ehrhart et al. 2014, use an instrumental variable approach to identify the causal impact of immigration. I use a shift-share instrumental variable approach, at the municipality level and country-municipality levels. This instrumental variable approach exploits the language commonalities between newly arriving immigrants and historic immigration settlements. For this approach to have plausible exogeneity, the necessary assumption is that earlier immigration settlements located in municipalities due to factors that do not affect current trade dynamics.

The results of this paper are as follows. First, I find that immigration is associated with an increase in exports at low levels of differentiation, and an increase in imports at low and high levels of differentiation. Both of these effects occur at the product extensive margin. Upon estimating a causal impact at a municipality level, I find that the main impact of immigration is through an increase in differentiated imports. In particular, an equivalent to doubling the average of overall migrant share to the population across municipalities would result in an increase in differentiated imports by 0.022%. This result is fairly consistent with the causal estimation at the country of origin and municipality levels, where I find an increase in differentiated and homogeneous imports at the extensive margin, and for imports at low to mid levels of differentiation as classified by Broda and Weinstein (2006)'s sector elasticity of substitution. Second, the presence of an impact on more differentiated imports, rather than exports, is suggestive that a preferences channel is more prominent than the fixed cost channel. Third, I show that there is significant heterogeneity in the impact of immigrants on trade, not only based on the level of product differentiation, but also depending on the country of origin characteristics. I find that immigrants arriving from countries with lower levels of ethnic fractionalisation and higher levels of interpersonal trust cause an increase in differentiated and homogeneous imports. Interestingly, immigrants from

less developed economies, by GDP levels, have a positive impact on differentiated imports, indicating that immigrants' specialised knowledge might be able to substitute weak institutions in their home country, a result that is consistent with some of the literature on the subject (Ehrhart et al., 2014)

My third chapter analyses the impacts of migration on the voting outcomes and electoral participation of Brazilian citizens in the 2018 elections. This election was marked by the rise of a far-right Presidential candidate, Jair Bolsonaro. My analysis focuses on the impact of one of the largest migration waves into Brazil, the Haitian immigration. Although the Haitian flow was first triggered by an exogenous push factor in Haiti, the Earthquake in 2010, and could be considered a quasi-natural experiment, the locations of migrants within Brazil have been far from random, especially after the unlimited issuance of humanitarian visas.

In an attempt to arrive to a causal estimate of the impact of the Haitian immigration, I employ a variation on the traditional shift-share instrument to predict the exogenous element in the location decisions of immigrants, which is similar is spirit to that employed in my second chapter. The difference in this chapter, however, is that I use the language proximity of Haitian Creole to other languages spoken by earlier immigrant groups to create a "Haitian unit equivalent" immigrant. The main assumption behind the validity of the exclusion restriction of this instrument is that historical immigrants located in Brazil due to factors that are independent from current political outcomes. I use this shift-share instrumental variable strategy, and discuss different robustness checks on its validity, to predict the number of Haitian immigrants received by each municipality. I conduct my analysis at the smallest administration unit in Brazil, the municipality.

I find that the Haitian immigration triggered a decline in voter turnout across different education groups, and led to an increase in vote shares towards the right-wing parties.[5] However, it seems to have had no impact on vote shares for Jair Bolsonaro or Fernando Haddad, the two leading contenders. In particular, a 1 percentage point change in Haitian immigrants ratio to the population causes a 0.97 percentage points reduction in turnout ratio, and an increase of 0.89 percentage points in vote shares towards right-wing parties, driven by an increase in vote shares towards the Presidential candidate Geraldo Alckmin, who was involved in an anti-Haitian immigration

[5]In my paper, I study the impact on aggregate categories of parties, right, centre, and left. For example, the right-wing parties is a grouping of multiple parties that are categorised as having a right-wing or centre-right ideology. I also present the impacts on the vote shares of specific candidates.

controversy. I find that the impact on political participation is present within smaller municipalities and ones characterised by higher than median Bolsa Família beneficiaries (i.e., more poor people). The increase in vote shares to Alckmin seems to be present within small-sized municipalities and those with lower levels of inequality. I analyse potential channels through which immigrants influence voting behaviour. I present suggestive evidence that the significant increase in vote shares to Alckmin is not solely driven by lower political participation, but also through potentially voters switching their preferences. I then proceed to study possible channels to explain the impact. I find little support for the idea that natives fear dilution of public resources, or for natives' flight away from municipalities that receive Haitian migrants; nor do I find evidence that immigrants lead to a shift in the types of candidates nominated by the parties. However, I present evidence suggestive of changes in voter behaviour occurring through cultural and economic mechanisms.

My findings are consistent with theories of political participation and voting behaviour in response to immigration. In addition, I find suggestive evidence that Haitian immigrants could be increasing the costs associated with voting or changing the electorate's priorities over their concerns, which potentially leads to higher absenteeism due to uncertainty regarding the candidate that best represents their interests.

The impact I found on the increase in vote shares towards right-wing Alckmin can be rationalised as follows. First, cultural related attitudes towards Haitian immigrants could be a potential explanation for the observed impacts. For instance, Alckmin was painted with an anti-immigration brush at the onset of the Haitian crisis. In his earlier capacity as governor of São Paulo, Alckmin got directly involved in a nation-wide political controversy against the arrival of approximately 800 Haitian immigrants from the flooded shelter in Acre at the start of the immigration (Cogo and Silva, 2016). Further, I find a significant reduction in vote shares towards Haddad and an increase in vote shares towards Bolsonaro within the sub-sample of municipalities where left-wing parties won in 2010, due to an increase in the Haitian immigration ratio to the population. This result could be indicative that the *native backlash* leading towards far-right parties was concentrated within the areas where the left-wing party historically dominated.

Through my analysis of cultural attitudes, I show that it is more likely that voters choose Alckmin or Bolsonaro over abstaining from the elections as the ratio of Haitian immigrants to the population increases. However, I find evidence that supports the *Contact Hypothesis* (Alesina and Tabellini, 2021) when studying the impact on vote

shares within municipalities where Haitian immigrants make up a larger share of the population. Within this sub-sample, I find that there is a reduction in vote shares towards Bolsonaro. These results lend support to the contact hypothesis, which states that exposure to a different cultural group can begin to create cohesion among the different groups that make up the society, and hence one would expect a reduction in vote shares to anti-migration candidates. This finding is consistent with other results in the literature (Steinmeyer, 2020).

I also find evidence for a positive economic impact on municipalities due to the presence of immigrants. An increase in the ratio of Haitian immigrants to the population led to positive impacts on value added in services and industry as well as the number of companies. Voters could be choosing the candidate that they view as most promising in delivering economic prosperity that would allow the assimilation of the new immigrant labour force. Based on the analysis of ideological positions across a multitude of topics (dataset based on Wiesehomeier, at al., 2021), Alckmin is a business-friendly figure who opposes all forms of state ownership of businesses and industry and promotes the deregulation of markets, more than Bolsonaro does (Wiesehomeier, at al., 2021).

Throughout this thesis, I make several contributions to the literature. The first chapter presents timely and extensive evidence on the impact of the COVID-19 pandemic on labour markets in Brazil. It thereby contributes to a rich and emerging literature documenting the impact of the pandemic on labour market outcomes. My paper presents an extensive analysis, one of the first to emerge, of the short-term and medium term effects of the pandemic shock within the Brazilian context – an emerging economy with a large informal sector and one that was still suffering from a recession in the years prior to the pandemic (Leyva and Urrutia, 2022; Prates and Barbosa, 2020). My work also exploits new waves of household survey data, including a special COVID-19 survey, created to document the unequal impacts of the pandemic. The data allow me to study, at the most granular level, how the heterogeneous government response across federal states was correlated with subsequent labour market outcomes, and how the response depended on worker characteristics (Petherick et al. 2020a, 2020b; Bouvier et al. 2022). Finally, the special COVID-19 monthly household survey data allows me to investigate the distributional effects and gendered impact of the government support program (Lara de Arruda et al. 2021; Lucia T Garcia, et al. 2021; Barbiera et al. 2021). To the best of my knowledge, my paper is one of the first to leverage the detailed household survey data to address pressing issues of inequality for policy makers in these ways.

The second and third chapters shed new light on the impacts of immigration, and in particular, South-South migration, as it relates to trade and political outcomes in a middle-income country. South-South flows have received relatively little attention in the literature compared to research on the impacts resulting from South-North or North-North migration. The context I study is unique in that the ratio of immigrants to the population is less than 1% in Brazil. However, it is perhaps salient that the receiving country attracted refugees from multiple new countries that were not part of the traditional immigrant profile in Brazil. In this setting, the effect of immigration on trade is not clear *ex ante*. My second chapter studies the universe of immigrants that includes voluntary as well as forced immigrants, and my third chapter studies only the impact of the forced Haitian immigration, which is one of the largest immigration episodes that Brazil witnessed.

In both of these chapters, I rely on novel data on immigration into Brazil curated and updated regularly by the Observatory of International Migration (OBMigra).

The second chapter adds to our understanding of the relationship between immigration and trade in a number of ways. First, I explore the impact of immigration on trade in the context of a middle income country and one with South-South immigration. Studying this context is important. Policies for the integration of migrants may be different in a country in the South (e.g. language courses, recognition of qualifications, employment assistance). These factors could affect the economic prospects of newcomers and their interaction with the host communities. My work also makes a methodological contribution: I construct a shift-share Bartik instrument based on the commonality of language spoken between historic immigrants and newly arriving ones. This novel approach allows me to arrive at a causal estimate of the impact of migration on trade and to overcome some of the limitations of previous papers that use an instrument based on the commonality of the country of origin. The latter is not adequate to be employed in a context like Brazil, because it received immigrants from many new countries which were historically not represented in the immigrant population. Empirically, my paper makes a contribution by studying the impact of immigration at the municipality level and at the origin-country-municipality level; I link immigrants to trade with their own country of origin. I present evidence that the impact of trade is heterogeneous and depends on the degree of differentiation of the traded product, at the extensive and intensive margins, and on the characteristics of the country of origin of the immigrant. Through the analysis of impact at different levels of product heterogeneity, I am able to point to some channels through which

immigration affects trade. A substantive finding is that immigrants from new countries or those with limited prior history of migration can also have significant pro-trade impacts.

The third chapter contributes to the literature on immigration and political outcomes in the following ways. First, I complement the academic debate in understanding how immigration impacts political behaviour in a middle income country and I shed light on the channels at play (Jaupart, 2018; Rozo and Vargas, 2021; Dustmann et al. 2019; Mayda et al. 2019; Tabellini 2019; Steinmayr, 2020). The discussion on what channels are the most important is evolving with new evidence, and can be context-dependent. Further, I study the context of Brazil, the fourth largest democracy in the world, at a unique moment in its history that marks the rise of a far-right presidential candidate after years of dominance by the left-wing party. The insights from my analysis may also provide implications for other emerging markets with a democratically elected populist leadership. My research includes evidence on the impact of migrants on electoral participation. This impact is mediated by the characteristics of voting communities: education, gender and age all matter. To the best of my knowledge, this effect has not yet been studied at such a level of disaggregation in the literature. Through this granularity, I examine how immigration can impact the voter composition and rationalise the findings within social and political theories of voter turnout (Solt, 2010; Kernet al., 2015; Dalton and Klingemann, 2011; Jost, 2021; Soss and Jacobs, 2009; Smith and Mickie, 2000). Understanding the effects on voter turnout and vote shares is important: voter turnout affects equal representation in a democratic system, and vote shares affect political outcomes and public policy reforms that have long-term welfare consequences. Although this goes beyond the scope of this paper, my future work will look into the consequences of changes in the electoral behaviour of natives on local social and economic outcomes.

My research serves as a useful benchmark for thinking about future episodes of immigration. Future migration episodes for Brazil are likely to occur as a consequence of climate change, political upheavals, and economic crises. Quantifying and rationalising the impacts of migrants on the host community is a key issue for policy makers. The new migrants, like the Haitians, are likely to arrive without sufficient economic means, may be traumatised by conflict, and face a high level of uncertainty regarding their mobility and length of stay in the host community. Hence, the impact that can arise due to the sudden arrival of a wave of unanticipated immigrants may be very different from that due to gradual, and voluntary immigration. Brazil is a country that had to

adapt its immigration laws to allow for the legalisation of humanitarian immigrants (Pacifico et al., 2015). The immigration structure and laws, alongside constrained fiscal resources, mean that the assimilation and effects of forced migration waves will be different in this context than that of a developed country. Finally, I expand the evidence on the channels of impact by providing descriptive analysis to assess cultural attitudes towards immigration as it pertains to political participation (Alesina and Tabellini, 2021; Tabellini, 2019).

Following this introduction, my thesis proceeds to present the three chapters that encapsulate the aforementioned research questions. The first chapter, tackling the labour market implications of COVID-19, is titled "The Impact of COVID-19 on Labour Markets and the Distributional Impacts of Auxílio Emergencial". This chapter is followed by two chapters titled "Immigration and Trade in a Middle-Income Country" and "Haitian Immigration and Citizens' Voting Behaviour',' that tackle the impact of immigration on trade and political outcomes, respectively. Each of these chapters is followed by an appendices chapter. Finally, chapter five concludes this thesis and draws political implications and future avenues for research arising from the analysis presented herein.

CHAPTER I

The Impact of COVID-19 on Labour Markets and the Distributional Impacts of Auxílio Emergencial

Abstract

This paper presents a short and medium term analysis of the impact of the COVID-19 pandemic on the Brazilian labour markets focusing on employment, wages and hours worked using the nationally representative household surveys PNAD-Continua and the new monthly data PNAD COVID. Brazil has a high prevalence of informality, a large number of small firms and a significant proportion of employment in contact-intensive sectors. Given this economic structure, this paper employs a conceptual framework to predict jobs at risk by sector and formality status, differentiating between demand and supply shocks. I then analyse the impact of the pandemic across different margins of heterogeneity related to workers' characteristics, and find that informal workers, females, youth and workers in small firms are the ones most adversely impacted in the short term. I document that sectors most susceptible to the shock because they are more contact-intensive and less teleworkable, such as construction, domestic services and hospitality, suffered large job losses and reductions in hours. In the medium term, I show that there seems to be a scarring effect on women's employment and wages, while informal workers seem to recover much faster than their formal peers. Given low income workers experienced the largest decline in earnings, extreme poverty and the Gini coefficient increased by around 9.2 and 5 percentage points, respectively, due to the immediate shock. I further study the role of the government response in mitigating

the shock of the pandemic and the distributional impacts under different counterfactual targeting scenarios of Auxílio Emergencial. I find that the government's broad based, temporary Auxílio Emergencial cash transfer program more than offset the labour income losses for the bottom four deciles, such that inequality and headcount poverty ratio decreased relative to the pre-COVID baselines.

The COVID-19 pandemic triggered major economic disruptions and severely affected labour markets around the world. Necessary policies to slow the spread of the disease and allow health systems to cope translated into an unprecedented economic shock. Containment measures abruptly reduced activity in many sectors and brought others to a complete halt. As a result, millions of workers lost their jobs and suffered income losses. Latin America was one of the worst affected regions, with employment contracting sharply, more than in other emerging markets (EMs) and advanced economies (AEs) (IMF, 2020). Brazil, in particular, ranked second in the world in the number of COVID cases by mid June 2020 (Veiga e Silva et al., 2020), and accounted for a quarter of the global deaths (Hallal, 2021).

In this paper, I focus on the Brazilian context and study how the COVID-19 pandemic affected labour markets in the short and medium terms. This paper tackles two themes in the context of the pandemic. First, I study the differential impact of COVID-19 on labour market outcomes depending on individual characteristics. Second, I analyse the role of governmental measures, such as the stringency of lockdown policies aiming to curb the health crisis, on assisting or impeding the recovery of the labour markets. I also focus on the distributional impacts of one of the most generous social support programs in Latin America, Auxílio Emergencial.

Brazil offered an interesting context in which to study the impacts of the pandemic due to the economic structure of the country on one hand, and due to the fragmented governmental policy response, on the other. Brazil was still recovering from the recession that gripped the country in 2014 when the pandemic hit. And unlike many advanced economies, Brazil has a high prevalence of informality (more than 45.9% of the labour force) and a majority of employed people in micro and small enterprises (51.6% of workers are working in firms with 1-10 employees). Those workers might be facing a larger probability of job losses than their counterparts in large firms potentially because the latter has a higher ability to draw on cash reserves or acquire external funding to retain employees (Bartik et al. 2020a; Apedo-Amah, et al., 2020).[1] Also workers in the informal sector tend to have occupations that are less fit for telework,

[1]Bartik et al. (2020a) find through a survey of more than 5,800 UK small businesses that many are financially fragile, and three quarters of the firms had cash to last them only two months or less. In contrast, a firm with monthly expenses over $10,000 had cash reserves to last it for only two weeks, while those with expenses less than $10,000, had one month of cash available. Apedo-Amah, et al. (2020) also show through a cross-country and a within country analysis of firms that smaller sized firms faced more severe financial constraints (both in terms of the available liquidity and fraction of firms in arrears) during the pandemic than larger firms. Guerrero-Amezaga et al. (2021), who use new survey data covering 35,000 small businesses in eight Latin American countries, also find that small firms were less aware of funding and aid programs than larger ones.

and thus will face larger labour income losses and potentially health concerns (due to a lower ability to self isolate if they work in essential sectors) when governmental containment measures are in place. Given this economic structure, losses to jobs and labour income can be more severe in a country like Brazil as opposed to other richer economies.

For the first theme in this study, I present an analysis of the impact on labour market outcomes within a conceptual framework developed by Alfaro et al. (2020) which has the advantage of taking into account the distribution of workers across firm sizes, occupations, sectors, and informality status. This analysis is done across three different phases of the evolution of the pandemic. Then the predictions from the framework are compared with actual labour market outcomes in the short term (Q2 2020). The methodology calculates the probability that an individual loses her job based on a sector-specific demand shock (due to fear of contagion), an aggregate demand shock, and a sector specific supply shock. The latter varies depending on occupation characteristics (contact-intensity and teleworkability), as well as on the firm's size. Then, I empirically investigate the heterogeneous impact of the pandemic depending on worker, sector and occupation characteristics in the early stages of the pandemic as well as in the medium term to shed light on recovery and scarring effects.

Applying the framework of Alfaro et al. (2020) and its adaptation by Komatsuzaki et al. (2020), I find that ex-ante employment-at-risk during the shut-down stage of the pandemic was close to 25 million jobs (approximately 27% of total employment in the 2nd quarter of 2019), with the concentration of job losses present in the trade, transport, hospitality and manufacturing (contact intensive and non-essential) sectors.[2] The demand component of the shock seems to have been responsible for more employment-at-risk than the supply shock, however, this result varies depending on the sector and income decile (for example, lower income deciles and some non-essential sectors like manufacturing face a higher impact from the supply component of the shock in the short term). As the containment measures are eased, the demand shock remains more persistent in its contribution to employment-at-risk while the impact of the supply shock dissipates. Comparing the estimates for ex-ante employment-at-risk with ex-post outcomes shows a strong correlation in terms of the sectoral distribution of job losses but realised losses overall were only about one-third of predicted employment-at-risk. A potential explanation of these differences pertains to Brazil's employment protection programs such as the formal employment retention program which had a

[2]Total employment in 2nd quarter of 2019 was approximately 94.2 million.

coverage of around 10 million workers, and the loan programs that assisted small and medium enterprises in financing working capital, payroll and in reducing the risk of default (Guerrero-Amezaga et al. 2021).[3]

I document, using the nationally representative household surveys PNAD Continua and PNAD COVID, that the recession induced by the pandemic presented new margins of adjustment when compared to the earlier 2014-2016 recession, and that the impact was highly unequal across individuals. Unlike its previous cushioning role in the earlier recession, informality shrunk in the short term impact of the pandemic. Also, there has been a much larger increase in discouraged workers and other people outside the labour force in comparison with the earlier recession. The unequal impact of the pandemic emerged across multiple margins of heterogeneity. Employment fell by over 18-34 percent year-on-year in vulnerable sectors such as construction, household services, and hospitality. The reduction in employment disproportionately affected informal workers, women, workers in small firms, and young cohorts of workers. Older cohorts of workers retained their employment but were more likely to be temporarily removed from their jobs, with high reductions in hours worked and monthly wages. Households in the bottom deciles of the income distribution were most affected, with extreme poverty and the Gini coefficient increasing sharply when measured at the height of the pandemic in May 2020. In the medium term, in the third quarter of 2021, women, youth and workers in small firms had persisting adverse impacts on labour income and hours worked. In contrast, informal workers were more negatively impacted than their formal counterparts, but I notice a stronger recovery in wages and weekly hours. I also notice a recovery in employment for the younger cohort of workers (between the ages 14-23 years old) after an initial significant reduction in employment at the onset of the crisis.

The second theme of my study analyses the role of governmental containment measures and the distributional impact of Auxílio Emergencial, Brazil's generous and robust social protection program that was implemented to combat the impact of the pandemic. Brazil had a fragmented response in terms of containment measures, with differing policies across municipalities and federation units, but the national government launched a comprehensive economic policy response. In terms of economic support, two key measures stand out (i) large-scale temporary cash transfers (Auxílio Emergencial – Emergency Aid – EA) to informal workers and poor households and

[3]Unfortunately, I lack the data to provide a direct assessment of the exact impact of this program, but it is an avenue for future research.

(ii) an employment retention scheme for formal jobs, which subsidised temporary reductions in hours. The former program had a cost of about 4% of GDP in 2020 and initially covered as much as 38.6% of all households (26.3 million out of 68 million) in Brazil, while the cost of the latter is estimated at around 0.6% of GDP. Around 25 percent of all formal private employment was covered by the employment retention program.[4]

I study the complex government response in two steps. First, using Oxford COVID-19 Government Response Tracker (OxCGRT), I analyse the impact of Brazil's response on the recovery of labour markets and sectoral employment in the subsequent quarter, focusing of the second and third quarters of 2021. Second, based on monthly survey data from PNAD COVID, I study the effects of Auxílio Emergencial on cushioning the impact of COVID-19 on low-income households. I find that contemporaneous lockdown measures led to a reduction in labour income by 0.23%. Reductions in employment were concentrated mainly in education and household workers, while lagged measures of lockdown stringency were positively correlated with employment, working hours, and sectoral employment in the subsequent quarter (3^{rd} quarter of 2021). With respect to Auxílio Emergencial (EA), I find that EA more than offset the negative impact on poverty and income inequality, at least temporarily. Inequality as measured by the Gini index would have increased from 0.53 pre-COVID to 0.58 post-COVID, not least because Brazil's large informal sector was deeply affected by the pandemic. Likewise, the poverty headcount ratio would have increased sharply from about 22% to 30%, corresponding to approximately 16.8 million individuals.[5] Instead, with the EA implementation, the poverty headcount ratio fell to 4.4 percent and the Gini coefficient to 0.51—both lower than their pre-COVID levels. Depending on the poverty line used, EA prevented between 10.5 and 20 million people from falling into poverty at the peak of the crisis. Throughout the analysis in this paper, I attempt to control for confounding factors that could be affecting the outcomes of interest studied. Despite the unexpected and sudden nature of the pandemic, causality is complex, and the results here are best interpreted as meaningful correlations.

Finally, in a counterfactual analysis based on survey data for October 2020, I simulate the re-distributive power of the EA under different targeting and benefit scenarios. I find that with better targeting, the same poverty reduction could have

[4]These numbers are based on the month of May, 2020.

[5]This analysis is based on the upper middle income poverty line of 5.5 USD per day. Based on the extreme poverty line of 1.9 USD per day, headcount poverty ratio would have shot up from 4.7% to 13.9%

been achieved at a lower cost. An amount just two-thirds of the observed transfer would have led to the same extreme poverty rate as the one observed in October, even after accounting for the administrative cost of better targeting. This shows the potential for a more cost-effective design should a renewed emergency cash transfer be implemented.

This paper contributes to a rich literature documenting the impact of the pandemic on labour market outcomes in the following ways. First, this paper presents an extensive analysis of this shock within the Brazilian context. In comparison to analysis done in advanced economies, Brazil is an emerging economy with a large informal sector, making policy options more complex. The economy was also, in the pre-pandemic period, still suffering from a difficult recession in late 2014 (Leyva and Urrutia, 2022; Prates and Barbosa, 2020). In this sense, my work is closest to the analysis by Bouvier et al. (2022) who study the individual transitions in the labour market as a result of the pandemic. Although my results are consistent with Bouvier et al. (2022), I take a different approach in my econometric analysis. In particular, I consider heterogeneous effects and incorporate the role of government policies in cushioning the impact of the shock. Second, my work relies on household survey data to document the differential and unequal impacts of the pandemic. In this respect, the paper is most closely related to the work by Piyapromdee and Spittal (2020); Benezeval et al. (2020), Adams-Prassl et al. (2020a, 2020b). Third, I study how the divergent government response across Federation Units was correlated with subsequent labour market outcomes, and conditioned by worker characteristics; for this analysis, I rely on a panel dataset of households across the surveys (Petherick et al. 2020a, 2020b; Bouvier et al. 2022). And fourth, by using the special COVID-19 monthly household survey data, my study also contributes to the literature by examining distributional effects and provides suggestive evidence of the gendered impact of the generous and wide-reaching government support program Auxílio Emergencial (Lara de Arruda et al. 2021; Lucia T Garcia, et al. 2021; Barbiera et al. 2021).

The remainder of the paper proceeds as follows. Section 2 summarises related strands of literature; section 4 describes the data and methodology used, and section 3 presents my results on a high-level overview of Brazil's labour market pre-COVID (section), and ex-ante assessment of employment-at-risk to the COVID-19 shock using the framework outlined in Alfaro et al.(2020) and compares it with ex-post labour outcomes (section 3.2). Section 3.3 reports empirical findings on the impact of COVID-19 across different margins of heterogeneity (formality, age, sector, gender). Section

4 analyses the impact of the government response of labour outcomes. Within this section, overall government stringency of lockdowns and economic support is analysed (section 4.1), followed by analysis of the the distributional impacts of the pandemic, the role of the EA in cushioning the shock for the most vulnerable, and counterfactual scenarios under various EA targeting options in section 4.2. Section 5 concludes.

1 Literature Review

This paper is related to a rapidly emerging empirical literature that documents the impact of the Covid-19 pandemic on labour markets and that documents and assesses the economic recovery as lockdowns are eased and the pandemic fades away. A number of papers have studied the impact of the pandemic within Brazil and the broader Latin American context. Veiga e Silva, et al. (2020) use public data from DATASUS (Department of Informatics of the Unified Healthcare System) (2010-2018) and the Brazilian Transparency Portal of Civil Registry (2019-2020) to test whether the actual number of infected individuals and deaths has been underestimated in official reports in Brazil. The authors find a significant difference between the actual numbers of infection and those being officially reported. Bouvier et al. (2022) use a panel dataset by merging quarterly PNAD Continua datasets in order to draw conclusions on the labour market flows due to the pandemic. They find that there is a significant increase in inactivity and a decrease in informality. They also document an impact of more than 11 million jobs lost in the second quarter of 2020, and two thirds of informal jobs destroyed.

In this paper, I employ an accounting framework developed by Alfaro et al. (2020) to calculate the ex-ante employment-at-risk across multiple sectors and to decompose the shock into its supply and demand components. Several other papers have used Alfaro's framework to construct similar estimates; e.g., Dingel and Neiman (2020), Koren and Petot (2020), and Leibovi et al. (2020). Barrot et al. (2020) focus on the sectoral impact of social distancing measures in France within a model of production networks and Baqaee and Fahri (2020) study the impact of supply and demand shocks in a multiple sector disaggregated model on employment, output and inflation. The Alfaro et al. (2020) methodology was used originally for Colombia. The appeal of using it in the Brazilian context is that the methodology considers jointly demand, supply and multiplier effects within an accounting framework that focuses on the role

of informal workers (a key characteristic of the Brazilian labour market) and differential firm sizes.

My work also presents evidence on the ex-ante characteristics of individuals that render them more vulnerable to the impacts of the crisis. In this sense, it is similar to work done by Cajner et al. 2020 and Coibion et al. 2020. Cajner et al. (2020) document the impacts of the COVID-19 pandemic on US labour markets using weekly payroll employment series. In particular, the authors find that cumulative losses in paid employment was estimated to be 18 million through April, twice the losses suffered during the Great Recession, with the hospitality and leisure sectors most affected, and the economy lost 13 million paid jobs between March 14 and March 28 alone. Similarly, for the US, Coibion et al. (2020) use a large-scale survey of households in the Nielsen Homescan panel to document impact on labour market characteristics in terms of employment losses. My analysis also provides evidence on the differential impact on self-employed individuals and small firms. This relates to work done by Humphries et al. 2020 and Bartik et al. 2020. My results show a similar impact of the pandemic to those observed in advanced economies in the aforementioned studies.

A recurring feature of my work is that it focuses on the heterogeneous impact of the pandemic on firms of different size. Guerrero-Amezaga et al. (2021) document, through a survey of 35,000 small firms in Latin America, that small and informal firms were less likely to be aware of the funding support available to them and perhaps in consequence applied less frequently and received less aid. In a similar vein, Humphries et al. (2020) show in the US context that small firms were less likely to know about the Paycheck Protection Program (PPP). Those that did apply were late to do so and were less likely to receive the aid. Further, in the short term, the authors find that approximately 60 % of small businesses laid off at least one worker during the pandemic. Around 37% of respondents reported that they were pessimistic about their ability to recover in the next two years (Humphries et al. 2020). This paper also complements the literature that studies the differential impact of COVID-19 on labour markets using survey microdata. In the case of Brazil, I use two large nationally representative household surveys, one of which is a special database released monthly to track the impact of COVID-19. In this respect, my work is closely related to Piyapromdee and Spittal (2020) who use data from the UK Household Longitudinal Study to document the differential labour market impacts of COVID-19 across workers incomes, characteristics and family structures. Benezeval et al. (2020) use new data from the Understanding Society: COVID 19 survey which is a special survey from the Understanding Society

panel collected over the past 10 years. The authors document the differential impact of the pandemic, showing that it affected mainly less educated workers without secure jobs. Adams-Prassl et al. (2020a, 2020b) use real time-survey evidence form the UK, US and Germany to show similarly unequal impacts of the pandemic on individuals. In particular, women, less educated workers and those working in jobs that are poorly suited for remote work are more likely to be affected by the crisis. In another study, the authors use three waves of data collected in March, April and May 2020 as part of the COVID Inequality Project in the US and the UK; from these data, they document wide variation in individuals' ability to work from home. This variation is found within as well as across occupations and industries. Further, the ability to work from home also differs based on worker characteristics. In general, highly educated employees reported a significantly higher share of jobs to be done at home.

Finally, I also contribute to the discussion that documents the heterogeneity of the impact based on existing inequality, employee and sector level characteristics such as the extent of contact-intensity and teleworkability (Blundell et al. 2020; Mongey and Weinberg 2020; Alon et al. 2020; Leibovici et al. 2020; Dingel and Neiman 2020). Alon et al. (2020) show that, unlike previous recessions, this pandemic affected negatively women's employment more than that of men's because women are more concentrated in sectors that are more likely to be subject to lockdowns and social distancing measures.

My work also relates to work on social safety nets in the times of crisis. Prates and Barbosa (2020) present a discussion on the emergency cash transfer program in Brazil, arguing that this program is essential but risks leaving out vulnerable formal sector employees. A more detailed analysis of Auxílio Emergencial is presented in Lara de Arruda et al. (2020).

Last but not least, this paper relates to the literature highlighting the importance of fiscal policy to tackle inequality and on the design of safety nets (Coady and Le, 2020) both in general and more specifically during COVID-19.

2 Data

The main sources of data are the PNAD-Continua household survey and the PNAD COVID special edition of the household survey.[6] Both surveys are representative at the 27 federation unit levels and at the national level, and are compiled by the Instituto Brasileiro de Geografia e Estatistica (IBGE). The PNAD-Continua is collected on a quarterly and annual basis and is available from the first quarter of 2012 to the third quarter of 2021. The survey reports indicators on income, hours worked, employment characteristics and individual level characteristics. For the analysis of PNAD-Continua, I use effective monthly income earned from all jobs. I change the nominal wages to real wages by using the survey's deflator, which also varies spatially. I exclude from my analysis households that indicate that they reported their income but their income is coded as missing in the survey (these constitute 0.08% of households for post-COVID income and 0.16% of households for pre-COVID income). The survey also includes information on whether the person is employed, the sector (at a four-digit level of dis-aggregation of the ISIC revision 4 classification) and occupation (following ISCO-08 classification), whether they are self-employed or salaried, different indicators on informality, size of the firm, education level of the person, among others.

PNAD COVID is a special and new edition of the PNAD-Continua survey and it is available from May to November 2020. This survey was conducted by telephone for approximately 48 thousand households per week and has two parts: the first one collects health indicators on COVID symptoms and hospitalisations; and the second one compiles data on labour market statistics including the employment status of individuals, their occupation and sector. The survey also provides detail on remote work (and whether that was paid or not), leave from work and reasons for the leave, and asks about income received outside work such as the EA related to COVID-19, Bolsa Familia, BPC-LOAS, retirement, unemployment insurance, rent and others.[7] The survey

[6]There has been a methodological update that was introduced in the 3rd quarter of 2021. The data included in this analysis is reflects the update published in February 2022 (accessed on 14th of June 2022). This ensures consistent historical comparison. That said, I have worked with the data prior to the methodological change and found the results in this paper to hold. I have also verified the main statistics reported relied on in this paper such as employment and unemployment rates, under-occupied individuals at the national and federation unit level match with the official estimates for the available quarters in the IBGE Automatic Recovery System - SIDRA. For more details on this update, I refer readers to Nota técnica 04/2021 Sobre a alteração do método de calibração dos fatores de expansão da PNAD Contínua..

[7]Bolsa Família is the Brazil flagship cash transfer program targeted to poor households with children introduced in the early 2000s and widely regarded as one of the driving factors for the substantial improvement of social indicators in Brazil during the 2000s. BPC-LOAS is the cash benefit to disabled persons.

also includes information on health outcomes, focusing on COVID-19 symptoms and methods of seeking health support. I consider income and hours "usually" received and worked to be the pre-COVID levels, and income and hours "actually" received and worked to be the post-COVID levels. In my analysis of poverty and the distributional impact of the EA, I only consider households that reported labour income. One important caveat is that the two surveys, PNAD-Continua and PNAD COVID, are not perfectly comparable. Therefore, whenever I combine the two surveys in the same analysis, I caveat the conclusions that can be drawn from it.

To compute the indirect supply and demand shocks in section 3.2 and equations (3.5) and (3.6), and in order to calculate the IO linkages, the analysis relies on the Brazilian 2019 IO matrix collected from the World Input-Output Database (WIOD). The IO matrix illustrates the flows between the sales and purchases (final and intermediate) of industry output and follows the National Accounts sector classification. In addition to these sources, I use the data on different classifications of containment measures provided by the COVID-19 Government response Tracker (OxCGRT) by Blavatnik School of Government at the University of Oxford. I also enrich my analysis with data from the System of Quarterly National Accounts and Google Mobility Reports.

3 Analysis of the Impact of the Pandemic on Labour Markets

3.1 Big Picture

The epidemiological situation in Brazil, in terms of number of COVID-19 cases and COVID-19 related deaths, was dire (as depicted in figure 4). Brazil ranked second in the world in the number of cases in mid 2020, and accounted for a quarter of the global deaths (Hallal, 2021). Alongside the introduction of lockdown measures, the pandemic had severe impacts on people's mobility, especially movement to and from workplaces, parks, and transit stations with a reduction range of approximately 30% to 50%. I depict the change in mobility during the pandemic to a pre-pandemic baseline in figure I.1(b). This figure uses a new dataset from Google that measures visitor numbers to specific categories of location (e.g. grocery stores; parks; train stations) every day and

compares this change relative to baseline day before the pandemic outbreak.[8]

In early 2020, Brazil's labour markets were still recovering from the 2015-16 recession. After reaching the highest level since the current unemployment series started in 2012 in early 2017, the unemployment rate only slowly fell from 13.7 percent to around 11 percent in late 2019. Looking at a broader measure of labour under-utilisation encompassing unemployed, underemployed and discouraged workers highlights the still high level of slack in the labour markets and steep reductions in gross saving rate until the 3rd of 2019 (Figure I.2(a) and I.2(b)). The COVID-19 shock thus hit with Brazil's labour market still in a fragile position. As in many countries around the world, job losses during the first months of the shock were unprecedented. Overall, as of August 2020, Brazil has lost roughly 12 million jobs relative to the same period in 2019, with most of the job losses concentrated in March, April and May. This far exceeds the employment destruction of around 4 million jobs during Brazil's deep 2015-16 recession. The vast majority of those who lost their job during the COVID-19 shock left the labour force, bringing the participation rate to a historic low of 55 percent. As a consequence, unemployment increased only modestly (Figure I.3).[9] In regional perspective, job losses in Brazil were somewhat less severe than in most peer countries. At the peak of the crisis in May 2020, job losses in Chile, Colombia and Mexico exceeded 20 percent of total pre-pandemic employment, close to double the losses for Brazil. The government's employment protection program likely played an important role here. It is worth pointing out though, that while employment had started to recover in peer countries by July 2020, Brazil was still experiencing additional job losses, bringing the impact closer in line with the regional average. The rolling quarter July-September saw the first job gains in Brazil after the initial shock, which continues in August-October and September-November. Job gains only erased a small fraction of the previous losses, and the number of unemployed workers continued to increase through the 3rd quarter of 2020.

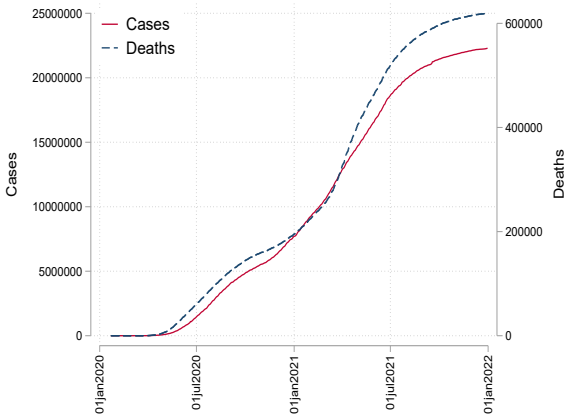
[8]Baseline days represent a normal value for that day of the week, given as median value over the five-week period from January3rd to February6th 2020. Measuring it relative to a normal value for that day of the week is helpful because people obviously often have different routines on weekends versus weekdays. Changes in movement do not take account of seasonal variation – for example, visitors to parks or outdoor spaces to be higher during the summer. The data may therefore reflect some changes in seasonal movements, rather than being fully explained by changes due to the pandemic. The amount of day-to-day variability in the raw data can make it difficult to understand how overall movements are changing over time. Therefore I convert the raw data into the rolling seven-day average.

[9]The sharply reduced participation rate following the COVID-19 shock is a common feature of Latin American labour markets but stands in contrast to advanced economies were unemployment increased more, likely because of stronger unemployment insurance (IMF, 2020).

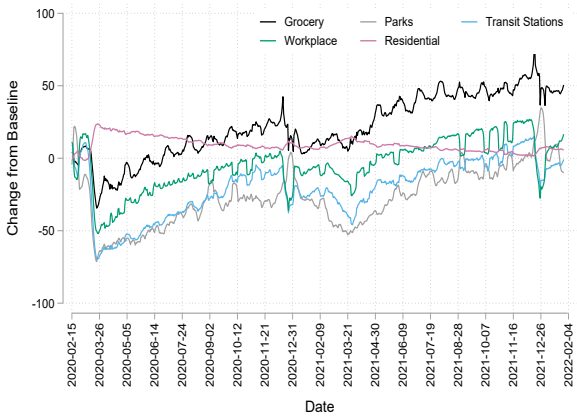
Interestingly, the labour markets seem to show signs of recovery. In the third quarter of 2021, I notice that the amount of discouraged workers and those in the potential labour force decreased from the quarter of last year. However, time related underemployment increased, masking the reduction in unemployment. However, gross saving rate increased by approximately 5.5 percentage points from the third quarter of 2020 to the third quarter of 2021. This is potentially driven by increase in savings by the poorest deciles that could be due to the generous Covid-19 social support programs and the reduction in spending during lockdowns. This is analogous to what have been documented in other countries like the United States. However, it might be expected to see these gains in savings dwindle with the retraction of the generous COVID-19 social transfers and sluggish income growth.

Figure I.1: Epidemiological Situation in Brazil

(a) COVID-19 Cases and Deaths



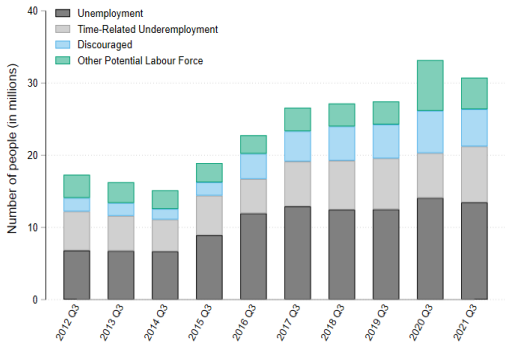
(b) Mobility in Brazil



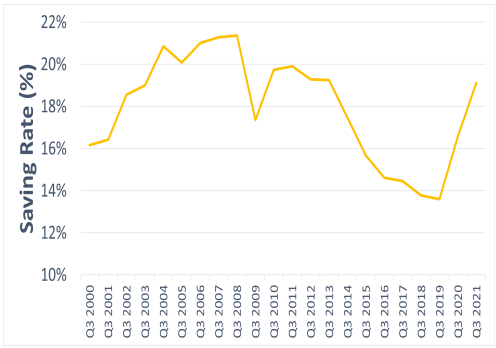
Notes: Graph I.1 shows the number of COVID-19 cases and COVID-19 related deaths from the beginning of the pandemic until January 2022. Source: Blavatnik School of Government – University of Oxford. Graph I.1(b) depicts changes in mobility from a baseline date prior to the onset of the pandemic. Mobility is depicted for five different categories: Grocery, Parks, Workplace, Residential and Transit Stations. Data is converted into a 7 day rolling average. Source: Google Mobility data.

Figure I.2: General Overview

(a) Unemployment, Underemployment, and Labour Force



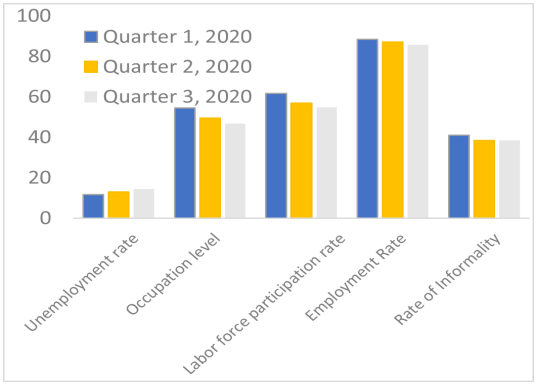
(b) Gross Saving Rate (%)



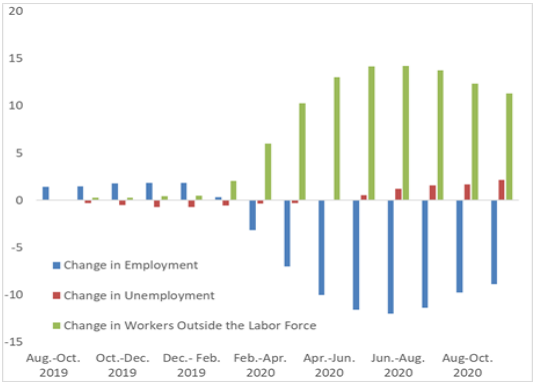
Notes: Graph I.2(a) depicts the number of people who are unemployed, are in time-related underemployment, are discouraged or are other members in the potential labour force. Time related Underemployed people are those who worked less than 40 hours a week but are willing and able to increase their working hours. It is a measure of labour under-utilisation. Potential labour force are those who are outside the labour force, and includes those who are discouraged and those who are seeking employment but can't find opportunities. Source: PNAD-Continua. Graph I.2(b) shows the saving rate (in %) for the third quarter of each year from 2000 till 2021. Saving rate is defined as gross savings divided by GNI. Source: System of Quarterly National Accounts.

Figure I.3: Key Labour Market Indicators

(a) Short Term Impact on First Three 2020 Quarters



(b) Quarterly Variations Relative to Previous Years



Notes: The first graph, figure I.3(a), shows the short-term impact of the pandemic on unemployment rate, occupation level, participation rate, employment rate and the rate of informality. Occupation level is the percentage of people employed in relation to people of working age. The second graph I.3(b), shows the changes in employment, unemployment, and workers outside the labour force by quarter. Source: Author's analysis using data from *O Instituto Brasileiro de Geografia e Estatistica (IBGE)* statistics.

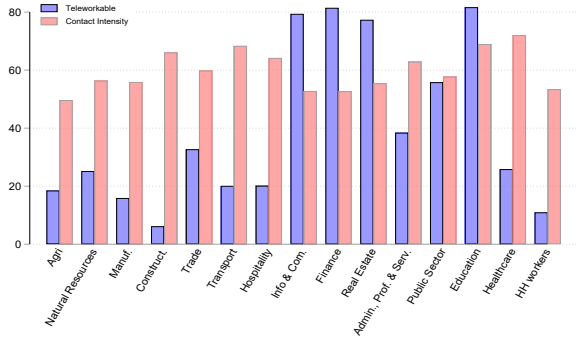
3.2 Analysis of employment-at-risk by Supply and Demand Shocks

To be able to look at labour market developments in Brazil in a more structured way, I turn to an ex-ante assessment of jobs at risk to the COVID-19 shock and compare this with the ex-post out-turns. To do so, I employ the framework outlined in Alfaro et al. (2020) which focuses on the initial period of impact of the crisis. In particular, I calculate the probability that individuals lose their job which depends on sector specific supply and demand shocks. The sector specific supply shock takes different values depending on the stage of the lockdown and the size of the firm, the characteristic of the worker's occupation in terms of its teleworkability and contact-intensity, the probability that this sector and its downstream and upstream sectors also face lockdowns and whether the person is formally or informally employed. I elaborate on the specific methodology in section 3.2.1.

This analysis assumes a differential impact of the pandemic depending on whether or not the worker is formally employed. Labour informality is the lack of compliance with labour regulations. However, there is a variety of labour regulations that govern workers and employers. Therefore, I define workers who are informally employed as those who do not have a formal contract (are either hired informally by formal firms or informal self employed workers), are not registered with the "National Register of Legal Entities", and do not contribute to the mandatory social security payment on the job. In the 2^{nd} quarter of 2019, informality was approximately 45.9% of workers.

The measure of teleworkability and contact intensity depends on the occupation of the worker and is based on Dingel and Neiman (2020)'s classifications. Dingel and Neiman (2020) classify 478 occupational categories in the American Community Survey, which follows the Standard Occupational Classification (SOC). Given the importance of contact-intensity and teleworkability, Figure I.4 presents the sectoral measures, taking the average of the teleworkability and contact intensity for the underlying occupations. In line with intuition, construction, domestic services, transportation and hospitality are especially vulnerable sectors given their low teleworkability and high contact-intensity as well as their higher share of informal workers.

Figure I.4: Teleworkability and Contact-Intensity Index by Sector



Notes: Estimates are based on my calculations. I aggregate the underlying indices of teleworkability and contact intensity at the sector level, and multiply the indices by 100. Higher values of either index imply that the sector is more teleworkable (can be done remotely or virtually) or more contact intensive (requires a certain level of contact with other people). Source: Dingel and Neiman (2020) and PNAD-Continua.

Table I.1: Labour Income and Hours Worked by Income Decile

	Average		**Median**		**Percent (%)**
	Labour Income	Hours Worked	Labour Income	Hours Worked	Informal Workers
Labour Income Quantiles					
10	273.8	25	300	20	9.3
20	681.2	33	700	32	8.7
30	997.5	39	998	40	3.7
40	1,066.3	39	1,000	40	4.5
50	1,269.5	41	1,260	40	2.8
60	1,514.7	41	1,500	40	3.6
70	1,886.8	41	1,950	40	3.4
80	2,366.9	42	2,500	40	3.3
90	3,459.0	42	3,400	40	3.4
100	9,779.4	43	7,000	40	2.8
Total	2,329.6	38	1,400	40	45.5

Notes: This table shows the mean and median labour income and hours worked per income decile. Deciles are created for all workers with a labour income based on an equal distribution of workers within each decile. Informal workers are those that do not have a formal contract (are either hired informally by formal firms or informal self employed), are not registered with the "National Register of Legal Entities", and do not contribute to the mandatory social security payment on the job. Source: PNAD-Continua.

3.2.1 Conceptual Framework to Analyse employment-at-risk

This section outlines the conceptual framework adopted in this analysis. I follow Alfaro et al. (2020) and I indicate where this analysis departs from the original framework. The analysis assesses the vulnerability of employment and income to the COVID-19 crisis and relies on data at the individual level prior to the crisis that contains relevant information on each worker's industry, occupation, firm size, and whether they are in formal or in informal employment. The risk to employment is assessed across the following dimensions: (i) Formality, (ii) Firm Size, (iii) Economic sector, and (iv) Occupation.

First, the link between risk and formality arises because of the flexibility that firms have in both hiring and terminating workers who are on informal contracts. This is because informal workers are not protected by any legal restrictions on firing and are not subject to employment retention schemes. Therefore, the impact on those workers could be expected to be large, relative to workers on formal contracts, in the immediate short-term after the crisis hits (Alfaro et al. 2020). However, for the same reasons, the rebound of informal workers can be faster than their formal counterparts. The rebound for informal workers may also be faster because many informal jobs require less physical capital, making them faster to adjust to conditions.

Second, firm size is an important determinant for the likelihood of job loss. Unlike small firms, medium and large firms can rely on their cash reserves and ability to access credit in order to retain employees for longer. So small firms will tend to face larger employment losses. In the recovery phase, larger firms can recover faster than smaller ones but both will be slower to recover than informal firms of comparable size.

Third, certain sectors in the economy are more subject to lockdown measures that were imposed by the government to contain the health crisis. Essential services were allowed to continue as well as sectors that were able to make the virtual transition (such as professional and internet services, including education).[10] Further, sectors that are more contact-intensive (i.e., the service provided requires high contact with people, as in the case of restaurants) have a higher likelihood of facing negative demand shocks.

Fourth and finally, the susceptibility to job losses also depends on the workers' occupation. In particular, occupations where the worker has to maintain in-person

[10]These distinctions are based on classifications of sectors in the United States; however in Brazil, it may be the case that preschool and primary education are poorly suited to virtual delivery. It also affects the ability of care-givers to go to back to contact-intensive jobs and limits online work.

interaction and cannot work remotely face a higher risk of unemployment and susceptibility to lockdown measures.

This analysis considers two types of shocks: *direct* demand and supply shocks and *indirect* demand and supply shocks. The direct demand shock occurs when demand falls specifically because consumers reduce demand out of fears for their safety; e.g., face-to-face interactions with consumers like hotels, theatres, restaurants, etc. The direct supply shock occurs when a sector faces a high probability of being in lockdown. These are mainly non-essential sectors and those where telework is not possible. However, these are not the only supply shocks that may arise. Voluntary closures of businesses and worker illnesses constitute additional supply shocks. This study focuses on the initial phases of the shock and so it doesn't yet account for the latter two shocks. The indirect supply shocks materialise when either the upstream or downstream sectors, or both, become affected by the lockdown measures. For example, if a lockdown measure prevents a company that produces certain intermediate goods from operating, then the firms using these goods in their final product become adversely affected. The indirect effect of the demand shock refers to the case when upstream sectors get affected by direct demand shocks; for example, when the food chain is affected by lower demand at restaurants. All the shocks described are further amplified by an "aggregate demand multiplier" resulting from shocks to income caused by the aforementioned channels which affects demand in all sectors. In what follows, I elaborate specifically on the framework and the components of the shocks.

The analysis incorporates three welfare measures: employment, labour income and value added per worker. I use the second quarter of 2019 as the baseline case before the COVID-19 pandemic and one that can be compared to the second quarter of subsequent years. The index i indicates individual, and employment is denoted by $(L_i = 1)$. Labour income is denoted by W_i and $F_i = 0, 1$ indicates whether individual i had a formal job at baseline. Let $Prisk_{st} \in \{Lrisk_{st}, Wrisk_{st}\}$; where $Lrisk_{st}$ is employment-at-risk and $Wrisk_{st}$ is labour income at risk. At risk scenarios are generated as follows:

$$Prisk_{st}^{formal} = \sum_{i \in s} \pi_{it} \times F_i \times L_i, \tag{3.1}$$

$$Prisk_{st}^{informal} = \sum_{i \in s} \pi_{it} \times (1 - F_i) \times L_i, \tag{3.2}$$

$$Prisk_{st}^{risk} = Prisk_{st}^{formal} + Prisk_{st}^{informal}, \tag{3.3}$$

$\pi_{it} \in [0, 1]$ is the probability of individual i losing their job due to the pandemic at time t. For value added in sector s, the equation is:

$$VArisk_{st} = Wrisk_{st} \times \left(\frac{VA^{IO}}{W^{IO}}\right)_s, \tag{3.4}$$

where $\left(\frac{VA^{IO}}{W^{IO}}\right)$ is the ratio of value added to labour income in s, obtained from the IO matrix.[11]

As mentioned earlier, each individual i is exposed to potential unemployment due to the crisis. However, this exposure depends on the sector and occupation characteristics of this individual. The probability that an individual i becomes unemployed and faces a reduction in income, $\pi_{it} \in [0, 1]$ is:

$$\pi_{it} = min\{1, (S_{shock_{it}} + D_{shock_s}) \times AD\}, \tag{3.5}$$

In this case, $S_{shock_{it}}$ is the supply shock affecting i's job. This parameter reflects any lockdown measures that force shutdowns in the worker's own sector or in upstream or downstream sectors; it also captures variation in whether person i is in an occupation that can be performed virtually. The supply shock takes the following form:

$$S_{shock_{it}} = S_{it} \times min\{1, (1 - T_i) + HighContact_i\} \times min\{1, Lock_s + L_{up_s} + L_{down_s}\}, \tag{3.6}$$

where T_i is the probability that individual i can perform their job virtually (telework) and where $HighContact_i$ is the probability that the worker's job requires a high degree of contact with other people. Both of these variables are calculated at the occupation level. An individual is not considered to be impacted by the supply shock if they can telework, even if their sector faces lockdown measures. The variable $Lock_s$ will take a value of zero if the sector will not be impacted by the lockdown measures, whereas sectors that are impacted to a certain degree by lockdown measures will have $Lock_s > 0$.

[11] Input and Output linkages tables (IO) are retrieved from World Input-Output Database for Brazil.

For the analysis in this paper, I rely on the Alfaro et al. (2020) classification of lockdown probabilities across sectors, and map the sectoral classification of Colombia in the original paper (the GEIH data) to that of Brazil, aggregating to one-digit sectors. The exact values of the lockdown impacts ($Lock_s$) are shown in table I.12 in the Appendix. The variables L_{up_s} and L_{down_s} represent the indirect effect of a lockdown through customers and providers, respectively. These reflect the possibility that, even if a sector is allowed to operate under lockdowns, it might still be impacted through the effects of lockdown measures on firms in other sectors that supply it with inputs and purchase its output. In particular:

$$L_{up_s} = \sum_{j \neq s} Lock_j \times \frac{purchases_{j:from:s}}{grossout_s}; \tag{3.7}$$

$$L_{down_s} = \sum_{j \neq s} Lock_j \times \frac{purchases_{s:from:j}}{grossout_s}, \tag{3.8}$$

Unlike the calibration in Alfaro et al. (2020), which assigns a value of either 0 or 1 for S_{it} for all firm sizes at a moment in time, I allow in this analysis for heterogeneity of the supply shock risk by firm size.[12] My approach is in line with the recent literature on this subject (ITC, 2020; Lambert and Van Reenen, 2021; OECD, 2020; Bartik et al. 2020a; Lin. et al., 2021, Komatsuzaki et al.,2020). The heterogeneity of closure risk by firm size stems from the differential ability of firms to retain employees in the short-term as well as in the longer term. Smaller firms, self-employed, and informal sector are assumed to lose employment faster when facing a lockdown. This is because they are not bound by firing restrictions or other employment protection regulations and have less cash reserves and access to credit (Alfaro et al. 2020). In contrast, larger firms would only lay off employees in case of a very protracted lockdown as these firms may be more able to transition to remote work, access credit more easily, and redirect resources within the firm to ensure higher employment stability (Bartik et al. 2020b; Lin. et al., 2021; Lambert and Van Reenen, 2021). The values of the calibration exercise are adapted from the study conducted by Lambert and Van Reenen (2021)

[12]Note that in this ex-ante analysis, I am abstracting from incorporating the cushioning effects of the government programs that were implemented. In section 3.2.2, I compare the estimates on employment from this exercise with actual employment changes and draw suggestive evidence on the role of the governmental programs implemented in Brazil.

using the BICS survey.[13] To calibrate the probability of firm closure in the immediate phase of the lockdown, I rely on the answers given by 8,764 registered businesses between 29th December 2020 and 10th January 2021 to the following question: "How much confidence does your business have that it will survive the next three months?".

The calibration also assumes that employment for small, self-employed, and informal workers recover more quickly when the economy reopens. This is because they face minimal organisation and physical capital as well as entry, hiring and firing costs. In contrast, due to fixed costs associated with, inter-alia, regulatory compliance, medium firms are assumed to have a harder time resuming operations when lockdowns are eased.

The variable D_{shock_s} is a demand shock affecting the sector of employment for person i, where the sector is indexed by s. AD is an aggregate demand shock as per equation (3.10). This shock is defined as a combination of a direct demand shock ($Direct_{Loss}$), weighted by the final sales to the sector's output ($FinalSales$) and an indirect demand shock reflecting weaker demand from buyers of s. The variable ($Indirect_{Loss_j}$) is constructed from the weights of purchases to total output ($purchases_{j:from:s}$). The exact values of the direct demand shocks are adopted from the consumption shocks to Sweden.[14]. The shocks are calculated as contractions in sales of consumer products prior to the implementation of lockdowns. These values are then applied to the sectoral specification used here. Table I.14 in the Appendix presents the exact values.

$$D_{shock_s} = Direct_{Loss_s} \times \frac{finalsales_s}{grossout_s} + \sum_j Direct_{Loss_j} \times \frac{purchases_{j:from:s}}{grossout_s}, \quad (3.9)$$

The aggregate demand shock AD is a standard Keynesian multiplier. [15] It is

[13]This is based on the BICS Wave 21, that was released on January 14, 2021 and is based on the prospects of UK firms. To the best of my knowledge, no equivalent survey was carried out for Brazilian firms, so I assumed that the probabilities of closure for firms is the same for those in Brazil. This is not ideal, as firms in Brazil may face different institutional and structural bottlenecks to their operation and closures.

[14]Sweden is chosen because in the initial stages of the pandemic, the country didn't implement any lockdown measures and so consumption shocks can be attributed to demand shocks. (This assumption requires abstracting from worker sickness, which might vary across sector.) Those are applied to Brazil through a set of cross-walks between the Swedish sectoral decomposition and that of Brazil. Exact calibration values are presented in table I.14 in the Appendix and rely on the calculations from Alfaro et al. (2020) that are based on the McKinsey 2020 Consumer Spending Survey.

[15]This number is based on heterogeneity in the marginal propensity to consume, which is higher for poorer households.

modelled as such:

$$AD = \sum_i \rho_i \frac{1}{1 - MPC_i} \tag{3.10}$$

This framework calculates employment-at-risk across different stages; however the specific methodology departs from Alfaro et al. (2020) methodology and follows the adaptation of the framework by Komatsuzaki et al. (2020). The first “shutdown stage” corresponds to the immediate impact in the second quarter of 2020. In this stage, non-essential sectors such as accommodation, arts and entertainment, hospitality, etc., are subject to lockdowns. The severity of the shock on workers depends on the degree of teleworkability and contact-intensity of the sector, as well as on inter-industry linkages. The shock from the demand side is also felt across multiple sectors, mainly transportation, storage, accommodation, hospitality, and services. To get a sense of how the labour market could develop after the immediate shock, two additional stages are analysed. During the second “partial opening stage,” lockdowns are eased for some sectors such as construction and mining. The key aspect here is the differentiation in resuming work across firm sizes. Specifically, in sectors where lockdowns are relaxed by 50 percent, it is assumed that workers in small firms, informal jobs, and self-employment can resume work with no restrictions. On the other hand, medium- and large-sized firms in the same sectors face more restrictions due to a higher composition of formal workers, higher costs of hiring and firing employees, and the difficulty of operating under social distancing protocols. These sectors are assumed to have their respective supply shock S_{it} decrease by half in the partial opening stage. Other sectors will continue to face lockdowns, which will translate into higher employment-at-risk as the prolonged lockdown will exhaust resources. Some firms might exit; hence the supply shock is assumed to increase by 1.5 times relative to the "shutdown stage". The shock from the demand side ($Direct_{Loss_s}$) is assumed to decrease by 25 percent relative to the "shutdown stage". During the final “advanced reopening stage”, those sectors that had partially reopened in the "partial opening stage" face no supply side restrictions, while those that were still facing lockdowns in the earlier stage, are assumed to have their supply shock S_{it} reduced to a quarter (one-fourth) of what it was in the "shutdown stage", to reflect a slow recovery due to potential financial stress from the prolonged impacts of the lockdown.

The demand shock in this stage ($Direct_{Loss_s}$) is reduced by 50 percent relative

to the shutdown stage as social distancing protocols are internalised by firms [16]. I present summary statistics on worker characteristics as well as framework parameters in tables I.2 and I.3. Table I.2 shows total employment, average labour income and hours worked, as well as the average idiosyncratic demand shock D_{shock_s} and the average lockdown probability $Lock_s$ by vulnerability to lockdown. Table I.3 presents the percentage of individuals with a certain characteristic, by less and more vulnerable sectors by probability of lockdown. In the Appendix, I also present the distribution of workers by formality status across different individual characteristics including their firm size, teleworkability of their occupation, and labour income relative to minimum wage in table I.16. I also show the share of employment by gender and education across measures of teleworkability and contact intensity in table I.15. To calculate individual job-loss probabilities, employment-at-risk is defined as the sum of all individuals employed prior to the pandemic weighted by their respective job-loss probabilities. The analysis does not incorporate furlough schemes and support programs (such as Brazil's emergency employment retention program).

Table I.2: Labour Market characteristics, Demand Shock and Probability of Lockdown by Vulnerability to Lockdown

	Total	**Average**				
	Employment (Trillions)	Hours Worked	Labour Income	Value Added by Worker (Thousands)	Idiosyncratic Demand Shock	Lockdown Probability
Vulnerability to Lockdown						
Less Vulnerable	40.59	36.9	2,864.35	63.293	0.034	0.035
More Vulnerable	53.56	38.8	1,828.39	40.937	0.118	0.755

Notes: This table shows total employment and average hours worked, income earned, value added, idiosyncratic demand shock (D_{shock_s}) and lockdown probability ($Lock_s$). Low Teleworkability and low Contact Intensity characterises all occupations with an index on each that is < 0.5. Vulnerability to lockdown is defined by $Lock_s$ and the idiosyncratic demand shock is defined by D_{shock_s} in equation (3.9). Sectors characterised as less vulnerable are those with $Lock_s < 0.5$ while those characterised as more vulnerable are those with $Lock_s \geq 0.5$. Source: PNAD-Continua and Dingel and Neiman (2020).

[16] Across the two phases, the exact values of lockdown across the stages are presented in tables I.12 and I.13.

Table I.3: Worker Characteristics and Job Attributes by Vulnerability to Lockdown

	Vulnerability to Lockdown		
	Less Vulnerable	**More Vulnerable**	**Total**
	%	%	%
Formality Status			
Formal	53.2	47.8	50.1
Informal	46.8	52.2	49.9
Firm Size			
1 to 5 people	51.5	50.3	50.8
6 to 10 people	9.1	10.1	9.7
11 to 50 people	12.3	12.6	12.4
51 or more people	27.2	27.0	27.0
Gender			
Male	52.4	60.5	57.0
Female	47.6	39.5	43.0
Education			
No education or less than 1 year of study	2.3	1.8	2.0
Incomplete elementary	17.9	25.7	22.3
Complete elementary	5.4	10.1	8.1
Incomplete high school	4.6	8.5	6.9
Complete high school	28.0	38.5	34.0
Incomplete higher education	8.0	5.3	6.5
Graduated	33.8	10.0	20.3
Contact Intensive Occupation			
Low Contact Intensity	65.0	60.9	62.7
High Contact Intensity	35.0	39.1	37.3
Ability to Telework			
Low Teleworkability	58.4	80.9	71.2
High Teleworkability	41.6	19.1	28.8

Notes: This table shows the percentage of workers employed in less or vulnerable occupations across different worker characteristics such as informality status, firm size, sex, education, contact intensive occupations and ability to telework. Low Teleworkability and low Contact Intensity characterizes all occupations with an index on each < 0.5. Sectors characterized as less vulnerable are those with $Lock_s < 0.5$ while those characterized as more vulnerable are those with $Lock_s \geq 0.5$. Source: PNAD-Continua and Dingel and Neiman (2020).

3.2.2 Results

Results from the Accounting Exercise This exercise uses the second quarter of 2019 PNAD-Continua data as a baseline. My results based on this exercise are as follows. First, I present the contributions of the supply, demand and inter-sectoral linkages shocks across the different stages of the pandemic. I find that ex-ante employment-at-risk in the shut-down stage was close to 25 million jobs, as shown in figure I.5(a). Employment-at-risk is reduced to approximately 14.5 and 6 million jobs respectively in the selective and advanced reopening stages, dropping by roughly 50 percent in the initial selective reopening phase and by another 70 percent in the advanced reopening phase. The 7 million employment-at-risk calculated in the advanced reopening state come from a somewhat protracted impact of the shock in sectors such as hospitality, transport and commerce, due to lingering demand shocks and the potentially longer recovery required for firms that already exhausted their financial resources. Looking at the reopening stages, even if quantitative estimates are necessarily imprecise, the general pattern is clear: a relatively fast rebound of a large share of jobs, coupled with lingering demand impacts in key sectors. A key question for Brazil will be whether the employment retention program can maintain employee-employer relationships permanently or the program simply delayed job losses rather than averting them. This is a question that can be investigated in future analysis when data on the data on the employment retention program becomes available.

Employment-at-risk is mainly concentrated in sectors that are less teleworkable and more contact-intensive, such as trade, transport and hospitality. The decomposition of the supply and demand shocks by sector in figure I.5(b) shows heterogeneity in the contribution of supply and demand shocks by sector in the immediate shock of the pandemic. In particular, my results show that the impact of the demand shock on employment-at-risk was higher than that of the supply shock in the hospitality, trade and transport sectors. I notice that in non-essential sectors such as manufacturing and construction, the impact of the supply shock seems to be larger than that of the demand shock, as the effect of lockdown measures might have been felt more in such sectors.

Figure I.6 shows employment-at-risk against wage-at-risk as well as value added per worker, by sector. Wage-at-risk is defined as the expected average wage loss per sector per worker. The first graph presents the wage at risk vis a vis the employment-at-risk for each sector, and the second graph shows the value added per worker (in logs) as

function of the calculated employment-at-risk by sector. I find that the average wage-at-risk is highest for those working in hospitality, transport and trade. This makes sense as these sectors rely on the interaction with other people and are subject to the implementation of lockdown measures. The transport sector, for example, might suffer a lot with restrictions imposed on people's movements. The trade, hospitality, and manufacturing sectors have the highest predicted employment-at-risk in the initial lockdown phase. Interestingly, I find that those sectors with the highest employment-at-risk during the immediate phase of the lockdown such as hospitality, trade, and transport, are not the ones with the highest value added per worker.

Second, I show that, for individuals who initially belonged to the lower income deciles, the impacts from the supply side of the pandemic (i.e., from lockdowns) were higher than those stemming from demand shocks. This result is consistent with findings based on the formality status of the employee. The accounting framework suggests that informal workers (who are more likely to belong to the lower income distribution) were impacted more by supply shocks than by demand shocks. This pattern contrasts with that experienced by formal workers. Informal workers faced almost double the employment risk of formal workers in the initial lockdown phase, because they worked in more contact-intensive sectors that were more subject to lockdowns. The data show that 77.5% of informal workers are in less teleworkable sectors as opposed to 64.4% of formal workers, see table I.16. Those results are presented in figures I.7.

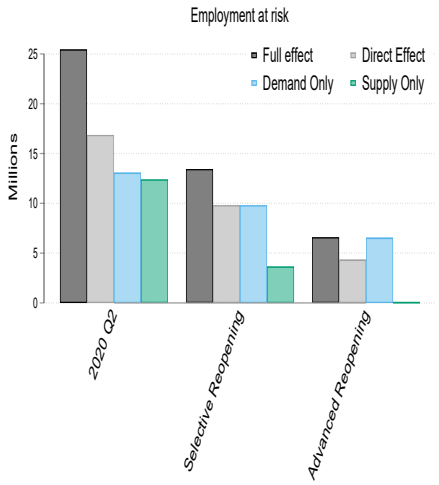
Third, I show in table I.7 the calculated employment-at-risk for workers at different firm sizes, by sector. employment-at-risk are highest among the self-employed and workers in small firms in the initial lockdown stage, as compared to workers in medium- and large-sized firms. The concentration of these losses are in sectors that are more contact-intensive and less teleworkable such as hospitality, transport and trade.

Comparing Framework Estimates with Actual Employment Changes Comparing the estimates of the ex-ante employment-at-risk with the ex-post losses in employment between the second quarter (Q2) of 2020 and second quarter of 2019 shows that the sectoral distribution of employment-at-risk is fairly well estimated with respect to actual changes in employment (figures I.9). This is also the case for formal and informal employment. and I.10). However, total job losses were significantly lower than the universe of employment-at-risk. In particular, while for large firms employment losses match predicted employment-at-risk, employment losses in small- and medium-sized firms (below 50 employees) were substantially below the level predicted

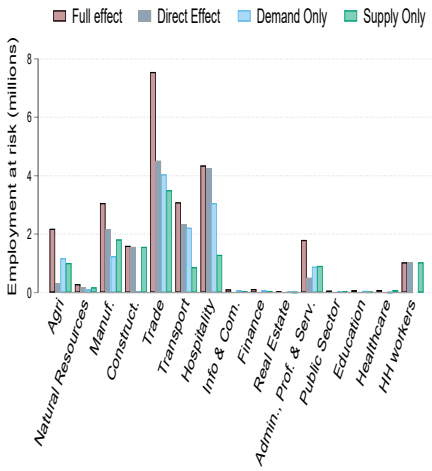
by the analysis. Several factors are at play here, but the government's employment retention program, Programa Emergencial de Manuten¸cão do Emprego e da Renda, which covered around 10 million formal workers, likely played an important role. This program allowed for a subsidised reduction in hours or temporary suspension of the employment relationship. Two other programs may also have led to differences between the model predictions and the observed outcomes. The Programa Emergencial de Suporte a Empregos provided loans for small and medium enterprises to cover their payroll. In the same spirit, state-owned banks offered subsidised credit to small and medium enterprises to assist with working capital. More generally, as many as 16.2 million workers were temporarily removed from work due to social distancing measures but officially remained employed (6.4 million of which unpaid).[17] Including these categories of workers (especially the unpaid ones) would bring "de facto" employment losses closer to the ex-ante employment-at-risk estimate. In the medium term of the crisis (2^{nd} quarter of 2021), employment-at-risk and actual employment losses are reduced for self-employed individuals and across all firm sizes (as shown in figure I.11).

Figure I.5: Overall Ex-ante Employment-at-Risk across Stages and by Sector

(a) Across Stages



(b) Across Sectors in the Immediate Stage

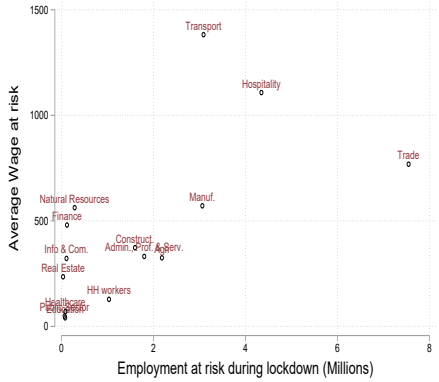


Notes: Figure I.5(a) shows the employment-at-risk in millions across three different stages: the immediate shock labelled 2020Q2, Selective Reopening and Advanced Reopening, further disaggregated into the impact of the full effect, the direct effect, the demand shock only and the supply shock only. Figure I.5(b) shows the employment-at-risk in millions across the different sectors, further disaggregated into the impact of the full effect, the direct effect, the demand shock only and the supply shock only.

[17]These numbers are based on the analysis of the PNAD COVID monthly survey in May 2020. See figure I.29.

Figure I.6: Employment and Wage-at-Risk in the Initial Lockdown Stage

(a) Employment-at-Risk and Wage-at-Risk across Sectors

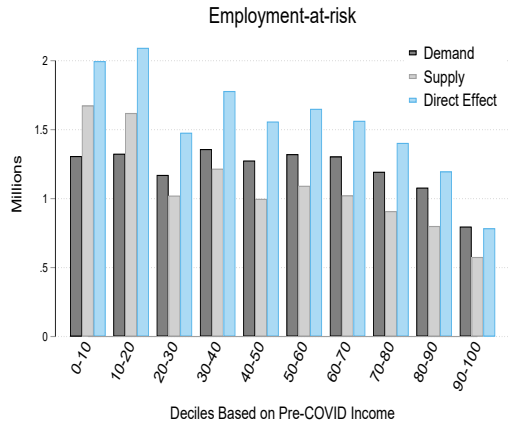


(b) Employment-at-Risk and Value Added by Sector

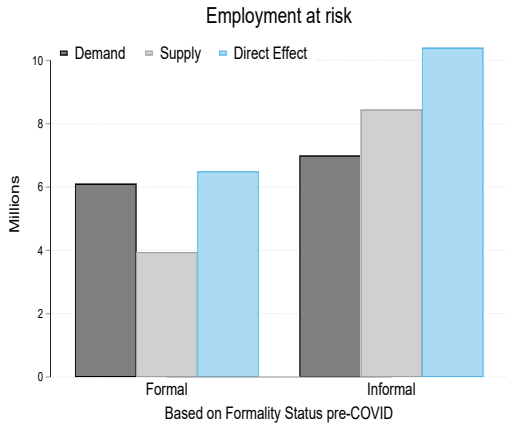
Notes: figure I.6 shows the average wage-at-risk plotted against employment-at-risk in millions. Figure I.6(b) depicts the value added per worker as a function of the employment-at-risk during lockdowns (the immediate impact) calculated based on the accounting exercise. Results are shown per sector.

Figure I.7: Decomposition of Shock Type by Income Deciles and Formality Status

(a) Employment-at-Risk By Income Deciles



(b) Employment-at-Risk By Formality Status



Notes: figure I.7(a) depicts the employment-at-risk during the initial phase of the lockdown (the immediate impact) calculated based on the accounting exercise, by decile and type of shock, while figure I.7(b) depicts the same but for formal and informal employment-at-risk.

Figure I.8: Ex-ante Employment-at-Risk

(a) Employment-at-Risk Across Sectors and Stages

(b) Employment-at-Risk Across Sectors by Formality

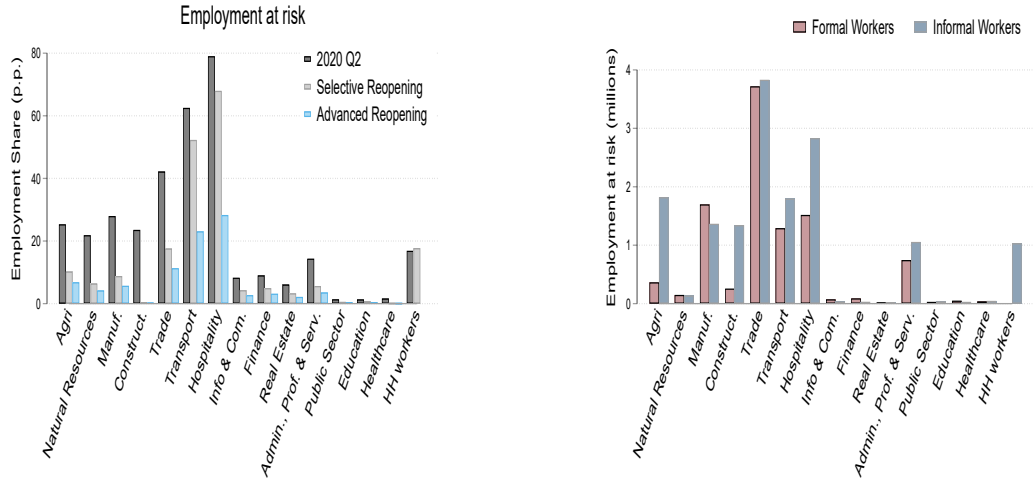


Figure I.8 shows the employment-at-risk in millions across three different stages: the immediate shock labelled 2020Q2, Selective Reopening and Advanced Reopening by sector. Figure I.8(b) shows the employment-at-risk in millions for formal and informal workers, across the same three different stages, disaggregated by sector.

Table I.4: Employment-at-Risk by Firm size (in millions)

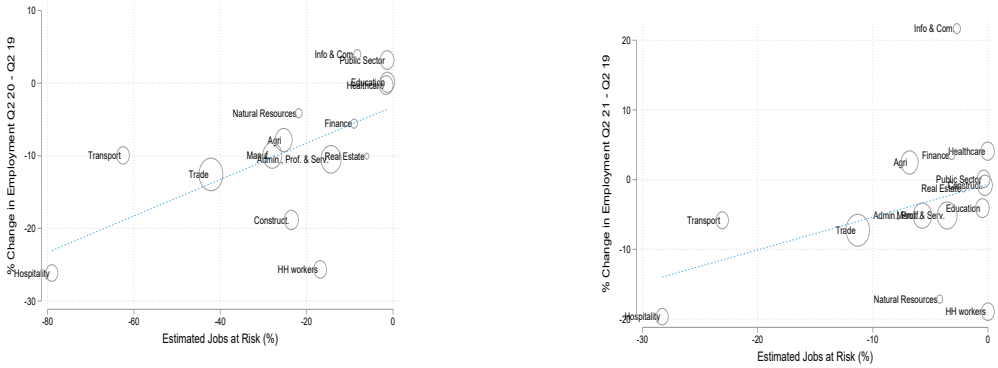
	Employment at risk (millions)			
	Self Employed	Small	Medium	Large
Industry				
Agri	0.588	1.332	0.132	0.132
Natural Resources	0.062	0.064	0.030	0.098
Manuf.	0.668	0.848	0.438	1.106
Construct.	0.852	0.537	0.095	0.113
Trade	1.718	3.702	1.038	1.087
Transport	1.331	0.530	0.234	0.900
Hospitality	1.027	2.397	0.641	0.281
Info & Com.	0.014	0.023	0.018	0.052
Finance	0.010	0.024	0.018	0.053
Real Estate	0.011	0.012	0.006	0.006
Admin., Prof. & Serv.	0.601	0.637	0.199	0.348
Public Sector	0.000	0.000	0.000	0.001
Education	0.003	0.007	0.011	0.016
Healthcare	0.014	0.025	0.010	0.020

Notes: This table shows employment-at-risk calculated for the immediate shutdown stage by industry and firm size. The numbers shown are in millions of workers.

Figure I.9: Ex-ante vs Ex-post Employment Changes

(a) Immediate Impact

(b) Advanced Reopening Stage versus 3rd Quarter of 2021

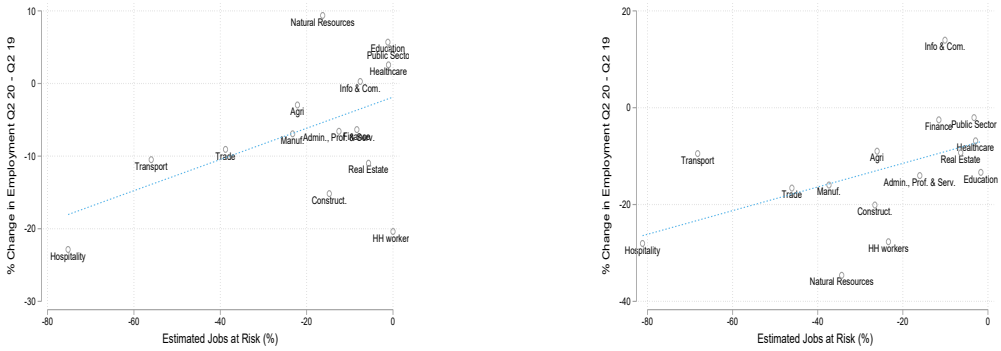


Notes: figure I.9 shows the actual percentage change in employment observed (between the third quarter of 2020 and third quarter of 2019) by sector as a function of the estimated employment-at-risk in the accounting exercise for the period immediately after the shock. Figure I.9(b) shows the actual percentage change in employment observed (between the third quarter of 2020 and third quarter of 2019) by firm size as a function of the estimated employment-at-risk in the accounting exercise for the period immediately after the shock.

Figure I.10: Ex-ante vs Ex-post Employment Changes

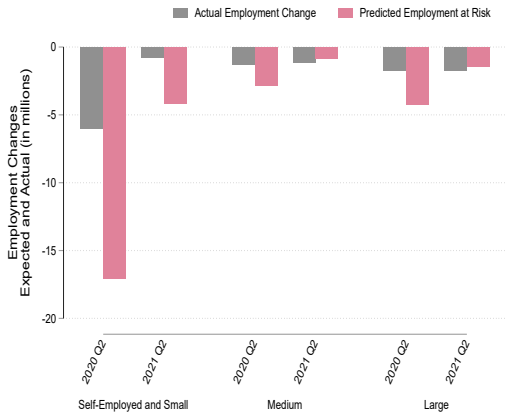
(a) Formal Employment

(b) Informal Employment



Notes: figure I.10 shows the actual percentage change in formal employment observed (between the second quarter of 2020 and second quarter of 2019) as a function of the estimated formal employment-at-risk in the accounting exercise for the period immediately after the shock. Figure I.10(b) shows the actual percentage change in informal employment observed (between the second quarter of 2020 and second quarter of 2019) as a function of the estimated informal employment-at-risk in the accounting exercise for the period immediately after the shock.

Figure I.11: Ex-ante vs Ex-post Employment Changes by Firm Size



Notes: The figure shows the actual percentage change in employment observed (between the third quarter of 2020 and third quarter of 2019) by firm size as a function of the estimated formal employment-at-risk in the accounting exercise for the period immediately after the shock.

In the following section, I turn to a more granular look at how the labour market shock affected different groups of workers, before turning to the analysis and distributional implications of the government response in the final section.

3.3 Heterogeneous Impact of the Pandemic on Labour Markets

Having shown the respective contributions of pandemic-related supply and demand shocks on employment losses, and having compared the predictions of the accounting exercises to actual changes in employment, in this section, I present stylised facts on the differential impact of the shocks on employment, hours worked, and income. These results can be disaggregated to show heterogeneous impacts that depend on a worker's labour formality status, age, gender, economic sector, occupation, and rural/urban location. Following this descriptive characterisation of the data, I draw together all the characteristics in a regression analysis that takes into account the impact in 2020 and 2021.[18]

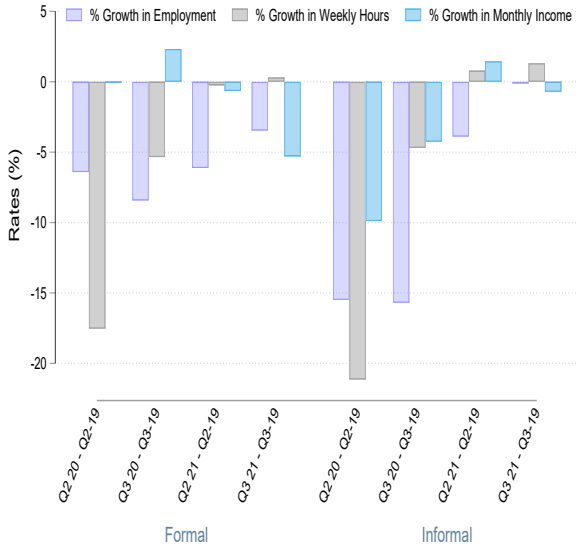
[18]Since the date of analysis, the third quarter of 2021 was the most recent quarter available for PNAD Continua.

3.3.1 Informality

In figure I.12, I present the percentage change in employment, weekly hours, and monthly income for formal and informal employees. The reduction in employment, income and hours worked was higher for informal employees and self-employed workers in the short term. However, in the 3rd quarter of 2021, I notice a more robust recovery in the labour market outcomes for informal employees than their formal counterparts. Interestingly, after observing preservation of monthly incomes for formal workers in 2020, monthly incomes seem to have decreased in the third quarter of 2021 with respect to the third quarter of 2020. These observations can be potentially explained by the implementation of the "Emergency Employment and Income Preservation Benefit" (BEM) program implemented in Brazil to support formal workers. The program, established in 2020, had the aim of preserving employment and income for formal workers who faced a temporary reduction in income or working hours, or a suspension in their employment. The program was later renewed in April, 2021 with a scheduled end date in August 2021, but was subsequently extended until October 24, 2021 (Brentani Toncolatto Advogados, 2021; Foccacia Amaral Lamonica, 2021). The reduction in incomes in the 3rd quarterof 2021 could reflect the termination of this emergency program for many employees who had been protected in the earlier temporary provisions of the law in 2020 or due to hiring decisions for lower incomes. could explain the large decrease in monthly income observed in the third quarter of 2021 . Using data from PNAD-Continua for Q3 2020 shows a reduction of approximately 11 million informal jobs relative to 3 million jobs lost for formal workers (Figure I.13) when compared with the same quarter of the previous year. As a consequence, the informality rate hit a record low during the peak of the crisis in 2020 Q2, falling by 6.1 percentage points relative to 2019 Q2, before reverting back to its pre-pandemic level in 2nd quarter of 2021 (as shown in figure I.14(a)). Figure I.14(a) also shows the reduction in the wage gap between formal and informal workers in the second quarter of 2020. The wage gap reached a ratio of approximately 47 percent from 53 percent in the same quarter of 2019. These patterns can be explained by the fact that many informal workers exited the labour force; formal sector workers, while being more likely to retain their jobs, faced sharp reduction in wages (as will be documented in later sections). However, the 2nd quarter of 2021 shows that the informal-formal wage gap is now even higher than it was in the same quarter in 2019. My analysis indicates that this is due to a larger reduction in the wages of formal workers (around 6.5%), and a small increase in

the wage of informal workers (0.9%). Using a logit regression analysis, I corroborate this observation by showing that informal workers were more likely to be doing remote unpaid work. This probability is almost three times as large for older cohort workers than it is for younger cohort workers (figure I.14(b)).

Figure I.12: % Change in Employment, hours worked and monthly income

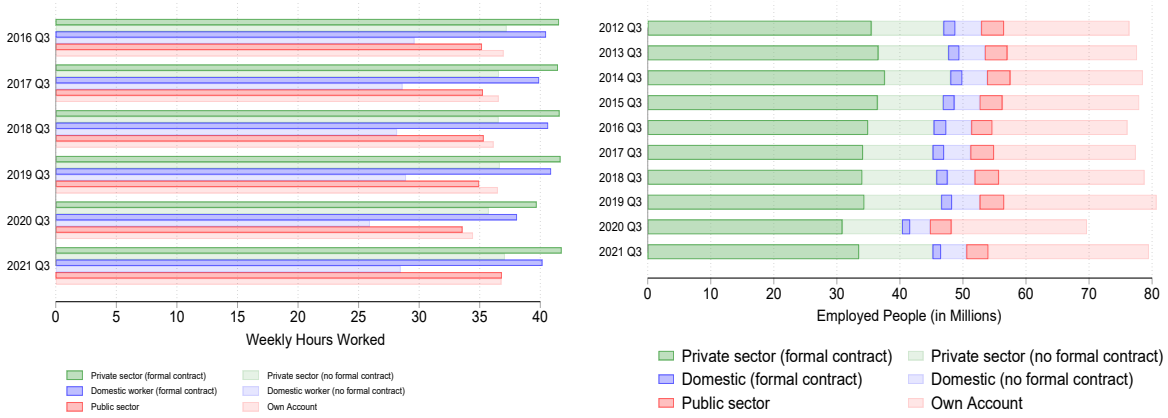


Notes: The graph depicts the evolution of the growth in employment, wage and hours worked by informality status.
Source: PNAD-Continua.

Figure I.13: Decomposition by Employment Categories

(a) Weekly Hours

(b) Total Employment

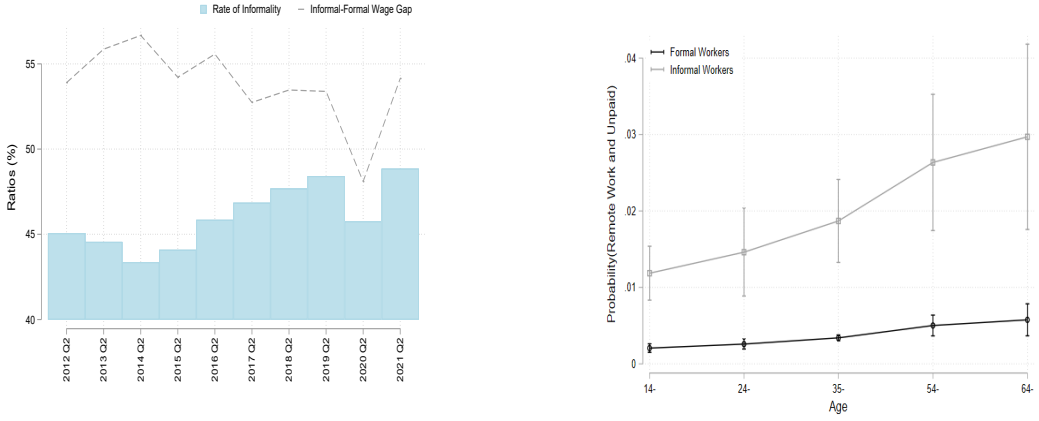


Notes: The figures show the number of hours worked (figure I.13(a)) and number of employed people in millions (I.13(b)) by type of employment. The data allows me to consider six types of employment: private sector employees (formal and informal), domestic workers (formal and informal), public sector employee, and own account workers. The estimates respresen the author's calculations of weekly hours worked and employment by broad groups of employment categories. Source: PNAD-Continua.

Figure I.14: Formal and Informal Work

(a) Informality Rate and Wage Gap

(b) Unpaid Remote Work Across Age Groups for Formal and Informal workers



Notes: figure I.14 depicts the evolution of the rate informality and and the informal-formal wage gap. The formal gap is defined as the ratio of informal workers to total employment. The informal-formal wage gap is calculated as the ratio between average monthly earnings for informal workers to that of formal workers at every point in time. Results are in percentages. Figure I.14(b) show results are from a logit model. Sampling design and weights are taken into account. Estimation includes federation unit fixed effects and industry fixed effects. 95 percent confidence interval shown. Source: PNAD-Continua and authors' analysis of PNAD COVID.

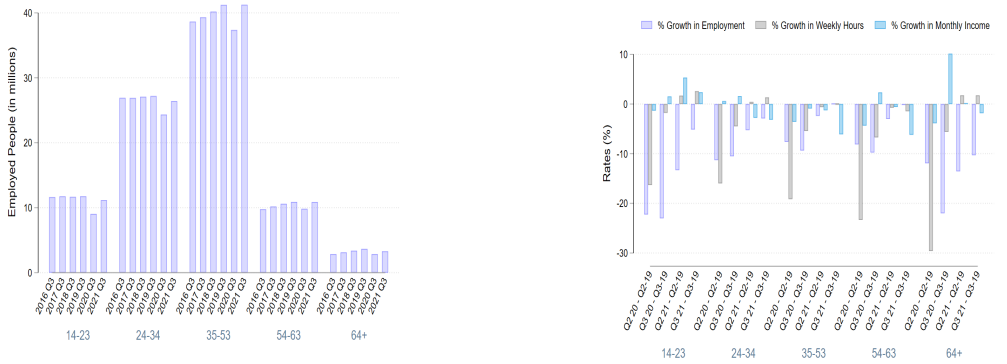
3.3.2 Age

Looking at patterns across age cohorts in figure I.15, I find that the younger cohorts faced the highest reductions in employment, while older cohorts suffered more pronounced reductions in hours on the back of milder decreases in employment. Older cohorts were more likely to be removed from work specifically due to social distancing measures, while retaining their employment status, resulting in an increase in income per hour worked during Q3 compared to a year earlier. This phenomenon most likely reflects the impact of the job retention program. In the 3rd quarter of 2021, there is a recovery in employment for cohorts below the age of 53, whereas the trend in employment reduction remained for older cohorts. This shock to older cohorts can have a scarring effect as this group of workers either chose retirement (from the supply side) or were laid off due to the hiring of more younger workers (demand side) at a lower wage than pre-pandemic as also shown in figure I.15(b).

Figure I.15: Decomposition by Age Categories

(a) Total Employment

(b) % Change in Employment, hours worked and monthly income



Notes: figure I.15 depicts the evolution of employment by age categories. Figure I.15(b) depicts the evolution of the growth in employment, wage and hours worked by age categories. Source: PNAD-Continua.

3.3.3 Gender

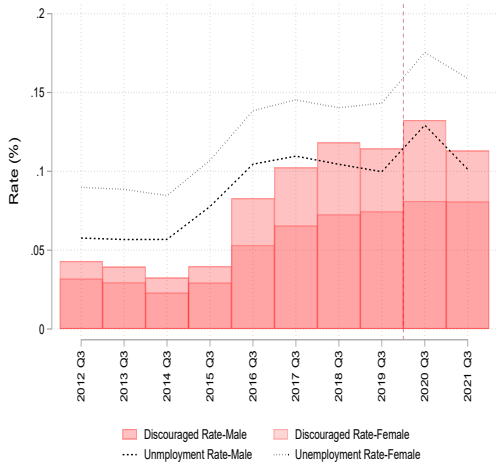
Historically, in Brazil, labour force participation and employment rates for females have tended to lag those for males by a considerable margin (Figure I.16). Between 2015 Q1 and 2020 Q1, the labour force participation rate and unemployment rate for women averaged about 52 and 13 percent, respectively, compared to 72 and 10 percent for men, resulting in a striking 20 percentage points wedge in occupational levels, which averaged 45 percent for women and 65 percent for men. I document a sharp decrease

across these rates in 2020 Q2 and Q3 due to the pandemic, to even lower levels than the rates prevailing during the 2015-2016 recession. In particular, the participation rate for women fell by about 6 percentage points to 46 percent in 2020 Q2, from 52 percent in 2020Q1 and continued decreasing in Q3. The rise in female unemployment rate would have been larger than what is observed (from 14.5 in 2020Q1 to about 17 percent in Q3) given that many workers left the labour force.

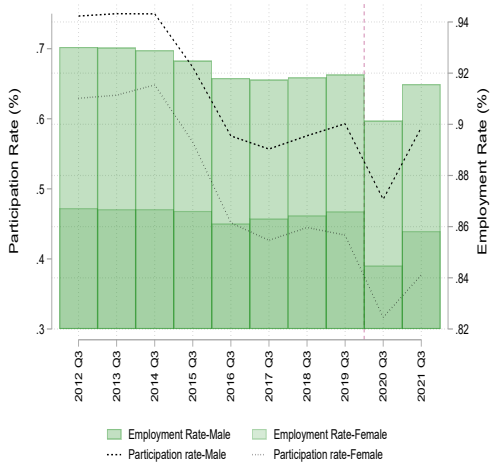
By comparison, the participation rate for men decreased by 5 percentage points to 66 percent in 2020 Q2 but rebounded slightly in Q3, while the unemployment rate increased from 10.4 percent in 2020 Q1 to about 13 percent in Q3. As shown in figure I.16, the proportion of both females and males exiting the labour force rose significantly in 2020 Q2, well above the levels observed in 2015 and 2016. However, while the flow out of the labour force continued increasing in Q3 for females, it slightly declined for males. Looking at percentage changes in employment, weekly hours and income per gender in figure I.17, female workers suffered a more pronounced fall in employment than males in both Q2 and Q3, and those who retained their jobs, experienced a more significant reduction in hours worked. In particular, hours worked by women decreased by 24 percent in 2020 Q2 compared to 2019 Q2 and rebounded marginally to -6 percent in Q3, compared to a contraction of 16 percent in Q2 and 4 percent in Q3 for males. Possibly because of the job retention program launched by the government and the sharper reduction in hours worked for women that retained their jobs, female workers experienced a slight increase in hourly income in 2020 Q3 compared to a year before. Interestingly, employment for males saw a recovery in the 2nd and 3rd of 2021, unlike employment levels for females, which did not recover to the pre-pandemic levels. Further, for both males and females, income did not recover and remained much lower than pre-pandemic levels, while weekly hours worked saw a recovery and even a noticeable increase for females when compared with pre-pandemic levels. From those results, the initial recovery phase is indicative of a scarring effect in employment and income for females (and in incomes for males). Women appear to be working more hours for less income, relative to before the pandemic.

Figure I.16: labour Market Impacts by Gender

(a) Discouraged and Unemployment Rates (%)

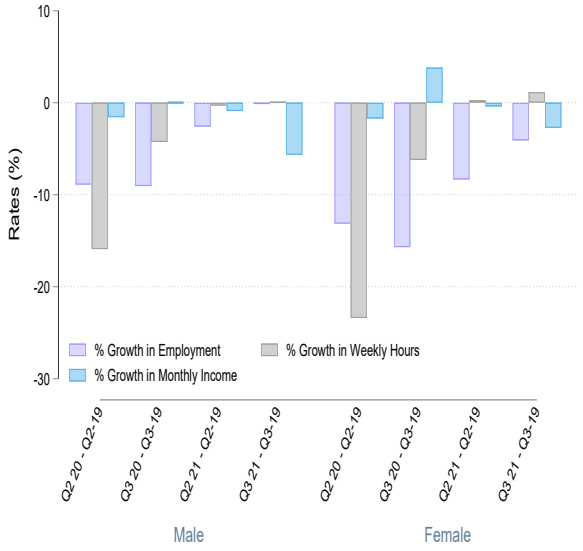


(b) Employment and Participation Rates (%)



Notes: The graphs depict the evolution of discouraged rates, unemployment, and employment and participation rates by sex. Source: PNAD-Continua.

Figure I.17: % Change in Employment, hours worked and monthly income



Notes: The graph depicts the evolution of the growth in employment, wage and hours worked by sex. Source: PNAD-Continua.

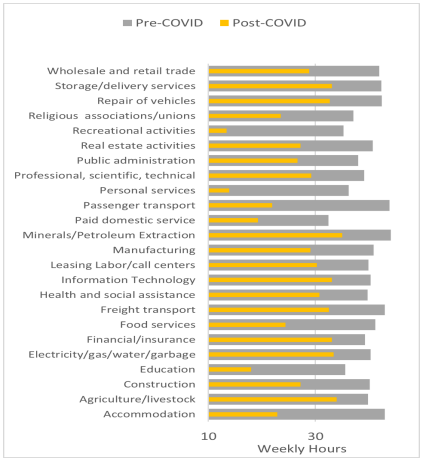
3.3.4 Sector and Occupation

The discussion of ex-ante employment-at-risk versus ex-post employment losses has already highlighted the sectoral heterogeneity in job losses. There is also considerable variation by industry in terms of changes in hours worked and labour income. On average, hours worked decreased from 39 hours per week pre-COVID to 27 post-COVID. In addition to having suffered the largest job losses, industries that are more contact-intensive – such as recreational activities, or personal and domestic services – experienced the largest drop in hours worked and income. For the month of May, females saw a reduction in monthly hours worked of 36 percent versus 27 percent for males. However, on average, income per hour worked for formal workers rose by 26 percent (28 percent for females and 12 percent for males). As noted, the government's employment protection program might have played an important role here, protecting incomes even as hours worked were reduced.[19] Using PNAD-Continua, I also document the impact of the pandemic on aggregate occupation groups as provided by the survey. Results are shown in figure I.19. I find that there is large heterogeneity in the impact across occupations, with lower-skill professions most adversely impacted in terms of employment. Those occupations saw a recovery in 2021 in terms of hours worked and employment, but monthly labour income remained less than its pre-pandemic levels. Directors also face a persisting negative growth in employment until the 3rd quarter of 2021, which can be attributed to hiring freezes and business closures. Another interesting category of workers is the "skilled agriculture" workers, who do not seem to have been impacted by the pandemic – perhaps due to the nature of their work (potentially intrinsically socially distanced) and to the work being viewed as essential activity.

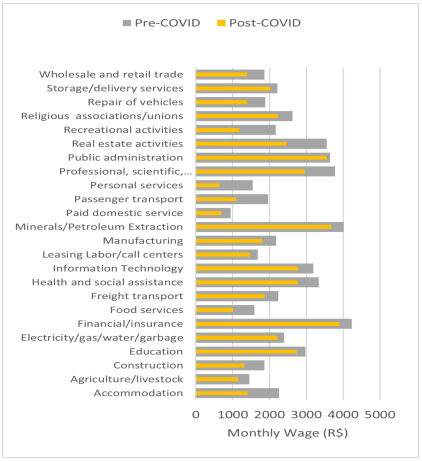
[19]PNAD COVID survey ceases in November 2020 so I cannot have comparable analysis on the recovery phase of 2021 using the same survey.

Figure I.18: Pre- and Post-COVID Hours Worked and Wages Across Sectors

(a) Weekly Hours Worked.

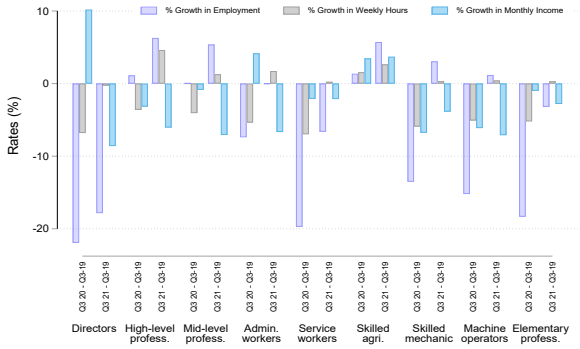


(b) Wages



Notes: figure I.18 depicts the author's calculations of weekly hours worked by sector for pre and post pandemic periods. Figure I.18(b) depicts the author's calculations of wages by sector for pre and post pandemic periods. Source: PNAD COVID (May, 2020).

Figure I.19: Pre- and Post-COVID Hours Worked and Wages Across Occupations



Notes: The graph depicts the evolution of the growth in employment, wage and hours worked by broad occupation categories. The presented occupation categories follow the Brazilian classification in PNAD-Continua. Source: PNAD-Continua.

3.3.5 Rural and Urban Regions

The reduction in employment was observed in rural and urban areas alike, reaching 12.4% in in 2020 Q3 relative to the same quarter of the previous year. In both urban and rural areas, those who remained employed saw little change in income but sharp reductions in hours worked. The fall in hours worked was noticeably sharper in urban areas (Figure I.20). Both areas saw a recovery of weekly hours worked, and in rural

areas, average weekly hours worked was even higher than pre-pandemic levels. What is interesting, is that unlike the urban area, rural areas saw a recovery in income and even an increase from pre-pandemic levels. This could be due to two things: First, rural areas had less prevalence of COVID-19. In the 2nd and 3rd quarters of 2021, COVID cases were still high in Brazil and were more concentrated in urban regions. Second, the increase in income could be driven by informal workers (higher concentration (17.8%) than formal workers (5.8%) in rural regions), as we have seen that the informality rate returned to its pre-pandemic levels or by the skilled agriculture workers as can be seen from figure I.19.[20]

Figure I.20: Employment, Weekly Hours, and Monthly Wage for Rural and Urban Regions

Notes: The graph depicts the evolution of the growth in employment, wage and hours worked by rural and urban regions. Source: PNAD-Continua.

3.3.6 The Importance of the Different Margins of Heterogeneity

Having looked at the differential impact of the pandemic across different dimensions and time periods, I summarise the patterns of heterogeneity of the COVID-19 impact on employment, wages and hours worked per week. I include one year prior to the pandemic, 2019, along with the second and third quarters and two post-pandemic years with the second and third quarters each.

In particular, I estimate the following equation:

$$Y_{it} = \alpha + \beta_1 X_{it} + \beta_2 X_{it} \times I_{2020} + \beta_3 X_{it} \times I_{2021} + \lambda_{UF} + \lambda_s + \epsilon_{it} \tag{3.11}$$

where Y_{it} are the dependent variables, namely: (i) whether the person is employed,

[20] But the large increase in income in rural regions cannot be all explained by informal workers' labour income, and further investigation is necessary.

which takes the value of 1 if the worker is considered employed and zero otherwise, (ii) whether the person is under-occupied, which takes the value of 1 if the worker is considered employed and zero otherwise (iii) monthly wage, and (iv) hours per week.[21] The vector X_{it} is a set of worker level characteristics: sex, education level, age, size of firm, extent of teleworkability and contact intensity of an industry, probability of lockdown, and race and high exposure to COVID-19 cases. with confirmed COVID-19 cases that exceed the 60^{th} percentile of the average cases of all states. I_{2020} and I_{2021} are dichotomous variables that take the value of 1 for 2020 (or 2021 respectively) and 0 otherwise. The variables λ_{UF} and λ_s are state and industry fixed effects respectively. Tables I.5 and I.6 show results on the main coefficients of interest (β_2 and β_3) on employment, wage and hours worked.[22] I estimate the probability of being employed or under-occupied with a logit regression based on the specification in equation (3.11), while the impact on wages and hours is estimated with ordinary least squares with individual characteristics, interacted with a dummy for 2020 to reflect the period immediately after the shock, and another for 2021 to reflect the longer-term impact, recovery, adjustment.[23][24] I find that most of the margins of heterogeneity discussed above mattered; the coefficients in the regression are individually and jointly significant, and there is no one single factor (say economic sector) that explains all the variation in the data. Consistent with the descriptive analysis in section 3.3, employment fell sharply for younger cohorts of workers, with those aged 14-23 being 1.39 percentage points and those aged 24-34 being 2.7% less likely to be employed than the older cohort (34+).

However, those who remained employed in the youngest cohort (14-34) worked more hours and earned more than the older cohorts, consistent with figure I.15. I also find that workers with secondary education were 1.51 percentage points less likely to be employed than the tertiary educated workers in 2020. However, the secondary and primary educated workers who remained employed in 2020 saw a statistically significant increase in their monthly wage and hours worked (in the case of primary educated workers) versus tertiary educated. A potential explanation for the latter effect is that

[21] Under-occupied in this definition refers to time related under-employment: Those are the individuals who worked less than 40 hours a week but are willing and able to increase their working hours.

[22] I exclude workers that report working zero hours and receiving a zero wage because the zero values might be conflating an actual income or working hours of zero or miss-reporting which would bias the results downwards.

[23] Each year has two quarters (2^{nd} and 3^{rd}).

[24] In relation to the conceptual framework, 2021 can be thought of as the year that represents the partial opening stage.

those primary-educated workers who retained their jobs were the highest-productivity in their skill category, while tertiary educated workers could have remained in employment but lost more in terms of labour income. Another possible explanation would relate to the effects of the job retention scheme for formal workers. I also find that, although females are not significantly less likely to be employed after COVID, those employed worked 1.11 percent fewer weekly hours. Black workers post-COVID were 1.39 percentage points less likely than white workers to be employed, but there is no significant difference in wage and hours worked for those who remained employed. Brown workers were not less likely to be employed post COVID vis-à-vis white workers, but those that remained employed witnessed an increase in their wage of 1.88 percent compared to white workers.[25]

[25]"Brown" is one of the racial categories used in the Brazilian data.

Table I.5: Heterogeneity of Impact by worker and industry characteristics

	(1)	(2)	(3)	(4)
	Employed	Under-occupied	Log(Wage)	Log(Hours per week)
Primary Educ X 2020	-0.00360	0.00598*	0.0912***	0.0465***
	(0.00380)	(0.00341)	(0.0136)	(0.00795)
Secondary Educ X 2020	-0.0151***	0.00425	0.0404***	0.00391
	(0.00366)	(0.00326)	(0.0114)	(0.00713)
Female X 2020	-0.000487	0.00251	0.00406	-0.0111**
	(0.00233)	(0.00214)	(0.00652)	(0.00536)
age 14-23 X 2020	-0.0139***	0.00726**	0.0311**	0.0525***
	(0.00329)	(0.00329)	(0.0122)	(0.01000)
age 24-34 X 2020	-0.0271***	0.0127***	-0.0189*	0.00279
	(0.00328)	(0.00320)	(0.0114)	(0.00918)
Black X 2020	-0.0139***	-0.00112	-0.0102	-0.000375
	(0.00462)	(0.00381)	(0.0112)	(0.00861)
Asian X 2020	0.0196	-0.0160	0.143**	-0.0189
	(0.0195)	(0.0163)	(0.0649)	(0.0377)
Brown X 2020	-0.00265	-0.000681	0.0188***	0.00592
	(0.00263)	(0.00239)	(0.00709)	(0.00595)
Indigenous X 2020	-0.0528***	0.0228	-0.0248	0.0417
	(0.0193)	(0.0141)	(0.0541)	(0.0318)
Pr. Lockdown X 2020		0.00303	-0.0297***	-0.0246***
		(0.00331)	(0.0102)	(0.00901)
Teleworkability X 2020		-0.000153	0.00789	-0.0245***
		(0.00297)	(0.00898)	(0.00736)
Contact Intensity X 2020		0.00671	-0.0216**	-0.0139**
		(0.00768)	(0.00953)	(0.00562)
<=10 employees X 2020		0.00850*	-0.0696***	-0.0588***
		(0.00449)	(0.00954)	(0.00538)
> 10 and <= 50 employees X 2020		-0.00497	-0.0485***	-0.0447***
		(0.00614)	(0.0102)	(0.00875)
Informal Worker X 2020		-0.0151**	0.00264	0.0335***
		(0.00715)	(0.0155)	(0.00963)
Informal Worker X Contact Intensive X 2020		-0.000590	-0.0212	-0.0331***
		(0.00807)	(0.0163)	(0.0110)
State F.E	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes

Table I.6: Continued: Heterogeneity of Impact by worker and industry characteristics

	(1) Employed	(2) Under-occupied	(3) Log(Wage)	(4) Log(Hours per week)
Primary Educ X 2021	-0.00329	0.00550*	0.0801***	0.00234
	(0.00365)	(0.00333)	(0.0119)	(0.00585)
Secondary Educ X 2021	-0.00619*	0.00425	0.0192*	-0.0124**
	(0.00356)	(0.00315)	(0.0103)	(0.00486)
Female X 2021	-0.0190***	0.00726***	0.0118**	0.00889**
	(0.00222)	(0.00197)	(0.00588)	(0.00381)
age 14-23 X 2021	0.0270***	-0.00461	0.0543***	0.0194***
	(0.00311)	(0.00294)	(0.0104)	(0.00597)
age 24-34 X 2021	0.000990	0.00371	-0.00684	-0.0140**
	(0.00305)	(0.00286)	(0.00971)	(0.00571)
Black X 2021	0.0101**	-0.00328	0.0159	0.0102
	(0.00432)	(0.00349)	(0.0100)	(0.00624)
Asian X 2021	0.0324*	-0.00700	0.0303	0.00642
	(0.0165)	(0.0133)	(0.0437)	(0.0223)
Brown X 2021	0.00260	0.000835	0.0202***	0.00164
	(0.00258)	(0.00226)	(0.00645)	(0.00379)
Indigenous X 2021	-0.00774	-0.00672	0.00195	-0.0420
	(0.0177)	(0.0126)	(0.0421)	(0.0453)
Pr. Lockdown X 2021		0.00677**	-0.0273***	-0.0152***
		(0.00313)	(0.00938)	(0.00545)
Teleworkability X 2021		0.00537*	-0.000868	-0.00956**
		(0.00288)	(0.00889)	(0.00487)
Contact Intensity X 2021		0.00333	-0.00227	-0.00646*
		(0.00815)	(0.00885)	(0.00373)
<=10 employees X 2021		0.000518	-0.0248***	-0.0113***
		(0.00414)	(0.00849)	(0.00383)
> 10 and <= 50 employees X 2021		-0.00303	-0.0195**	-0.0108***
		(0.00566)	(0.00950)	(0.00415)
Informal Worker X 2021		-0.00735	0.0736***	0.0298***
		(0.00795)	(0.0137)	(0.00642)
Informal Worker X Contact Intensive X 2021		0.00272	-0.0473***	-0.00462
		(0.00843)	(0.0148)	(0.00740)
State F.E	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes

Notes: The table shows regression results based on the specification in equation (4.14) and PNAD-Continua micro-data. Standard errors are in parenthesis. Survey data design and sampling weights are taken into consideration in the estimation. Controls include sex, formality status, education level, age, size of firm, extent of teleworkability and contact-intensity of an industry, probability of lockdown, and race. The omitted categories are as follows: on education it is tertiary education, on age it is 54+ cohort, on race it is white, and on firm size it >50 employees. Robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Workers in sectors that were subject to a higher probability of lockdown worked 2.46 percent fewer hours and earned 2.97 percent lower monthly wages. Workers in sectors that are highly contact-intensive worked 1.39 percent less hours and earned 2.16 percent less in terms of their wages. Workers in teleworkable sectors worked 2.45 percent less hours and earned 0.07 percent more than those in less teleworkable ones (albeit not statistically significant). Workers in small (<10 employees) and medium-sized (between 10 and 50 employees) firms earned 6.9 percent and 4.8 percent lower monthly wages respectively while working 5.88 percent and 4.47 percent less, respectively, than workers in large firms (above 50 employees). Workers in firms with less than 10 employees were also more likely to be under-occupied. Informal workers worked 1.44 percent more hours than formal workers and were less likely to be under-occupied than their formal peers. However, those informal workers who were employed in highly contact intensive sectors were working 3.31 percent fewer hours than informal workers employed in less contact-intensive sectors in 2020.

Looking at the medium-term impact, I find that in 2021, females remained less likely to be employed than males and significantly more likely to be under-occupied if employed. There seems to be some recovery in employment prospects for younger cohorts of workers (14-23 years old), in terms of an increase in monthly wage, hours worked and employment relative to their older counterparts. The increase in hours worked observed in this study is consistent with the lower likelihood for this group of being under-occupied. Recovery is also noticeable for black and brown individuals. Workers in sectors with a high probability of lockdown remain more likely to be under-occupied, earn lower wages and work fewer hours, with magnitudes similar to those observed in the immediate phase of the shock in 2020. Recovery in hours, employment and wages (statistically significant on the latter) is seen for workers in both contact-intensive and teleworkable occupations. Results show that informal workers in 2021 were earning more and working more hours than their formal-sector peers, except for those who were employed in contact-intensive sectors, who remained at a wage disadvantage compared to other informal workers in 2021. This result is consistent with the conjecture made in the conceptual framework: informal workers, due to lower start-up costs, seem better able to bounce back after the pandemic than their formal counterparts.

In summary, the labour market picture in the medium term is complex. Women and those workers in sectors with high lockdown probability and high contact intensity tended to remain under-occupied and unemployed in 2021. This can be considered as

evidence of a scarring effect for this group of workers. Furthermore, there is a decline in under-occupied status for workers in small firms (fewer than 10 employees) and medium-sized firms (between 10 and 50 employees), those who are in employment were earning lower wages and working fewer hours than their peers working at firms with more than 50 employees.

4 The Impact of the Government's Response to the Pandemic

In this section, I turn my attention to the analysis of the impact of the government response towards the pandemic on labour market outcomes, with a focus on the generous social support program, Auxílio Emergencial. This analysis complements the previous sections of this paper in that it sheds light on how the unprecedented government response, intended to curb the health crisis, could have also resulted in unequal labour market outcomes across individuals.

Up until late 2020, Brazil remained one of the most affected countries in the world by COVID-19, based on epidemiological indicators of spread, mortality and ability to impose social distancing measures (Veiga e Silva, 2020). In the initial stages of the pandemic, the federal government denied the severity of the virus and the importance of vaccines. The Bolsonaro administration rejected international recommendations for non-pharmacological interventions, promoted misinformation about the virus, and lacked a national order that would govern lockdown measures and mask-wearing. Despite allocating R$ 42 billion (USD 7.8 billion) to exclusively address the ramifications from the COVID-19 crisis, the government response was not enough to curb the rise in cases or high mortality - especially among pregnant women and indigenous population. Data on the spread of COVID-19 is complicated by the high prevalence of under-reporting, due in part to limited testing. These epidemiological facts highlight the inadequate health response of the federal government towards the pandemic (Ferigato et al., 2020).

Thus, Brazil was characterised by a heterogeneous and fragmented response at the municipality and the Federation Unit (UF) levels. This fact motivates the question for this section: what was the impact of the heterogeneous pandemic response across federation units on the labour market outcomes of individuals? To answer this question,

I make use of this varied federal response and elaborate on the methodology in what follows.

This section is divided into two parts. First, in section 4.1, I analyse the impact on employment, hours worked and labour income arising from (i) strictness of the lockdown policies and (ii) the economic support provided. I also estimate the impact on sectoral employment due to those measures. Next, in section 4.2, I focus on one of the major forms of economic support, that of the generous social support program implemented due to the pandemic, Auxílio Emergencial. I estimate the impact that this program had on poverty and inequality and I provide different counterfactual scenarios on its administration and distribution. I elaborate on the methodology and results in what follows.

4.1 Containment Measures and Economic Support

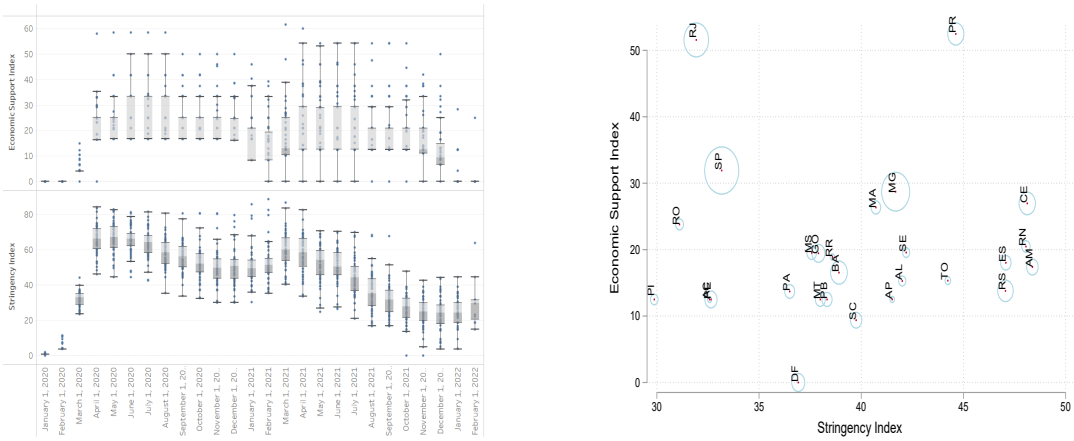
I begin this analysis by studying if and how the lagged and concurrent (based on the third quarter of 2021) government lockdown policies, on one hand, and economic support measures, on the other, could have assisted or impeded the recovery of labour market outcomes of individuals in Brazil across Federation Units. To provide insights on the recovery, I focus on the most recent quarters of 2021. In order to conduct this analysis, this section uses the Oxford COVID-19 Government Response Tracker (OxCGRT) at the Brazilian sub-national level, the Federation Unit (UF), for proxies of the government response. This dataset provides updated and comparable information on policy measures on a daily basis. I provide descriptive analysis per Federation Unit on the different measures reported in the study for Brazil, such as restrictions on public gatherings, school and workplace openings, public events, contact tracing and other interventions to contain the spread of the virus. I present the summary statistics for all such measures in the Appendix in tables I.28, I.29, I.30, I.31.

The tables show the heterogeneous response across the 27 UF regions. For example, Rondonia had the highest Stringency Index (representing strictness in lockdowns) while Espirito Santo had the highest level of economic support. For the analysis on the impact of such containment measures, I focus on two measures: Stringency Index and Economic Support Index. The Stringency Index reports the strictness of lockdown restrictions that govern people's behaviour, including measures taken to promote public information campaigns. The Economic Support Index measures income support

and debt relief.[26] In the case of Brazil, the Economic Support Index would capture the implementation of the "Programa Emergencial de Manuten¸cão do Emprego e da Renda" aimed at formal employees, "Programa Emergencial de Acesso a Crédito (PEAC)" aimed at reducing the risk of default for SMEs, "Programa Emergencial de Suporte a Empregos" aimed at providing loans for SMEs to cover payroll for up to 4 months, and a program by "Caixa Econômica Federal and Sebrae" that provides loans from state-owned banks to SMEs to cover their working capital (Guerrero-Amezaga et al., 2021). The Economic Support Index would also capture Auxílio Emergencial, aimed at informal workers and unemployed individuals. The evolution of each measure and its inter-quartile range are shown in figure I.21(a), and the heterogeneity across Federation Units is documented in figure I.21(b).

Figure I.21: Stringency and Economic Support Indices

(a) Evolution and Distribution of the Indices (b) Value of the Indices Across Federation Regions



Notes: figure I.21(a) depicts the interquartile range of the Economic Support and Stringency Indices across the 27 Federation Units in Brazil and over time. Figure I.21(b) depicts the values of the Economic Support and Stringency Indices across the 27 Federation Units in Brazil in the third quarter of 2021. Source: The Oxford COVID-19 Government Response Tracker (OxCGRT).

I assess the impact of the government response in two ways. First, I analyse the impact of each of the measures (Stringency Index and Economic Support Index) on employment, weekly hours worked and monthly income. Second, I analyse the heterogeneous impact of the two measures based on individual characteristics such as education level, sex, age and race. For this analysis, I make use of the PNAD-Continua survey design in order to match employees across the second and third quarter of 2021.

[26]For specific information on the construction of each of those variables, I refer readers to the index methodology elaborated on in the GitHub repository for OxCGRT.

The focus on these two quarters is interesting because it allows me to see the impact of containment measures in the second quarter of 2021, when COVID-19 cases where still high, on labour market outcomes in the third quarter of 2021, when there were decreases in the number of cases as well as the stringency of lockdowns. Hence this analysis aims to provide suggestive evidence on the role of governmental policies in the recovery of the labour markets.

Matching individuals across the survey datasets will allow me to control for individual characteristics and lagged outcomes variables in an attempt to minimise many of the endogeneity concerns that surround this context. This leaves me with a sample of approximately 89.47 million employees. Due to the rotational nature of the survey, not all workers are matched.[27] I find that the matched sample differs from the unmatched one of employed individuals across many dimensions when compared using a simple difference in means and t-test p-value. However, those estimates can be highly sensitive to the (large) sample size. To attenuate this issue, I follow Imbens and Wooldridge (2009) and report the normalised differences. Based on this approach, I find that the two samples actually do not differ significantly in the observable characteristics of the workers. I report the results of the balance test in table III.32.

Impact of Government Measures on Labour Market Outcomes First, and in order to estimate the impact of the Stringency Index and Economic Support Index on labour income and weekly hours, I estimate equation (4.13). Imposing lockdown measures (captured by the Stringency Index) or providing economic support is highly correlated with the spread of the virus, and in turn correlates with mobility, wages, and hours worked. However, establishing causality in this context is difficult. In particular, the estimation is susceptible to reverse causality, and this will create biased estimates. As a way to attenuate this bias, I focus on the period where COVID-19 cases started to decrease very steadily reaching their lowest levels since the beginning

[27]My matching methodology follows an adaptation of "the basic matching" approach that was first proposed by Lopes (2002) (Bouvier et al. 2022). This approach relies on a creating an individual identification key which is composed of, among others, the household identification number, birth date, gender, race, and UF region. This approach is rather strict and will miss many individuals, potentially due to measurement error and misreporting. Both aspects are likely to represent precarious attrition and represent less than 5% of the sample (those who do not report a birth date). Actual attrition due mobility or refusal to be re-interviewed, as Bouvier et al. (2022) point out, may be more of an issue when matching pre- with post COVID quarters. However, I only match two post COVID quarters to maximise the matched sample size and present balance tests between the matched and unmatched individuals. i also use the most recent updated data of PNAD-Continua that revisited ts weighting structure to reflect, among other things, the attrition issue that occurred in the data collection after the pandemic.

of the pandemic (as of the time of this analysis in the 3rd quarter of 2021).[28] This specification controls not only for contemporaneous values but also for lagged values of the number of COVID cases and COVID related deaths, as well as lagged values of the government measure under study (e.g., the Stringency Index or the Economic Support Index).

Since I match individuals across surveys, I am able to include lagged values of the economic variable studied $Y_{i,t-1}$. Because there is large variation across occupations and sectors in terms of the ability to telework and avoid the shock of lockdown measures or sick workers (both supply shocks), and across firm sizes in their ability to retain employees and draw on cash reserves to support income, I include occupation, sector and firm fixed effects. Further, as can be seen from the results in table I.7, the pandemic has a disproportionate impact depending on worker characteristics such as age, sex or race. I also include fixed effects on each of the aforementioned individual characteristics.

$$E_{i,T} = \mu_1 + \sum_{c}^{C}\sum_{h=0}^{3} Beta_h C_{c,UF,T-h} + \sum_{h=0}^{3}\phi_h Cases_{UF,T-h} + \sum_{h=0}^{3}\kappa_h Deaths_{UF,T-h} \tag{4.12}$$

$$+ \lambda_{education} + \lambda_{sex} + \lambda_{race} + \lambda_{age} + \epsilon_{iT}.$$

$$Y_{i,T} = \mu_1 + \sum_{c}^{C}\sum_{h=0}^{3} Beta_h C_{c,UF,T-h} + \sum_{h=0}^{3}\phi_h Cases_{UF,T-h} + \sum_{h=0}^{3}\kappa_h Deaths_{UF,T-h} + \omega_2 Y_{i,T-1} \tag{4.13}$$

$$+ \lambda_{education} + \lambda_{formality} + \lambda_{firmsize} + \lambda_{occupation} + \lambda_{sex} + \lambda_{race} + \lambda_{age} + \epsilon_{iT}.$$

In this specification, $E_{i,T}$ is a binary variable indicating if a person is employed or not, employed in informal work or not and considered a discouraged person or not, where T $\in$ 3rd of 2021. $Y_{i,T}$ represents the weekly hours worked and income earned (in logs) in the 3rd of 2021 across all individuals in the sample. $Y_{i,T-1}$ represents the lagged value (by one quarter) of the outcome variables. $c \in C_{c,UF,T-h} =$

[28] This is also the period that allows me the most efficient merge with the PNAD Continua in the second quarter of 2020 and one that preserves the largest matched sample size.

$\{\text{Stringency}_{UF,T-h}, \text{Economic Support}_{UF,T-h}\}$ so $\sum_c^C \sum_{h=0}^3 C_{c,UF,T-h}$ represent the Stringency Index or Economic Support Index (contemporaneous and lagged). The lagged values of the index are important because they control for the path dependency in policy implementation and for the different Federation Unit institutional prowess. Because labour market outcomes vary at the quarter level, I aggregate the daily measures into this level as well.[29] The summations $\sum_{h=0}^3 Cases_{UF,T-h}$ and $\sum_{h=0}^3 Deaths_{UF,T-h}$ represent lagged and contemporaneous COVID cases and COVID related deaths aggregated at the quarter level. The benefit of the aggregated indices at the quarterly level is that the average incorporates the variation across the entire quarter. $\lambda_{education}, \lambda_{formality}, \lambda_{firmsize}, \lambda_{occupation}, \lambda_{sex}, \lambda_{race}, \lambda_{age}$ are education, formality status, firm size, occupation, sex, race and age fixed effects. Therefore, the variation that I am exploiting is within those characteristics of the workers, but across federation units that have different levels of the stringency and Economic Support Index.

I present the results (coefficient $Beta_0$) in table I.7. This estimate compares the impact of the economic support index or stringency measure on workers who resided in UF regions with higher values of these indices, versus those that resided in regions with lower values of these indices, controlling for firm size, occupation, sex, education, race and age group. I find that the contemporaneous Economic Support Index is correlated with lower probability of employment (0.34 less likely to be employed if an individual lives in an area with higher levels of Economic Support Index), but is not correlated, in a statistically significant way, with hours worked or income earned. The contemporaneous measure of the Stringency Index is also correlated with a lower probability of being employed as well as with lower income earned (higher levels of governmental lockdown measures are correlated with a reduction in labour income by 0.23 %). The result on the latter is less surprising, considering that the Stringency Index is capturing the degree of which lockdown measures are implemented and thus the ability of firms in susceptible sectors to operate. The negative coefficient on the Economic Support Index could be reflecting reverse causality, whereby economic support will be higher in regions where people suffered more employment losses. This economic support index is reflecting the emergency cash transfer program (Aux'ilio Emergencial) that was directed at informal or unemployed individuals. The positive coefficient on the lagged values of the Stringency Index for probability of employment, hours worked and income could be reflecting the importance of stringency measures

[29] I take the average of the Stringency Index and the Economic Support Index and the summation of the COVID-19 cases and deaths.

(reflecting lockdown measures) in curbing the impact of the pandemic contemporaneously and thus allowing firms to resume business in the next quarter. The lagged Economic Support Index is positively correlated with probability of being employed in the subsequent quarter, indicating a potential employment recovery for employees who in previous quarters where in areas that received higher economic support, but at reduced hours worked (potentially indicative of under-occupied employment).

In the second part of this analysis, I disentangle the heterogeneity of the impact of each of those indices on hours worked and income earned for further insights. This latter results could be due to the employee retention scheme implemented in Brazil for formal workers.

Table I.7: Impact on labour Income and Hours Worked

	(1) Employed	(2) Informally Employed	(3) Discouraged	(4) Log Diff in Hours	(5) Log Diff in Income
Economic Support Index	-0.00314***	0.00345***	0.00168***	0.000319	-0.000602
	(0.000457)	(0.000693)	(0.000357)	(0.000282)	(0.000640)
Stringency Index	-0.00404***	0.000246	0.00157***	-0.000132	-0.000592
	(0.000491)	(0.000782)	(0.000364)	(0.000319)	(0.000729)
L1 Stringency Index	0.00338***	0.00170**	-0.00204***	0.000626*	0.000570
	(0.000518)	(0.000827)	(0.000383)	(0.000340)	(0.000779)
L2 Stringency Index	-0.00304***	-0.00564***	-0.00244***	-0.000581	-0.000212
	(0.000675)	(0.00106)	(0.000510)	(0.000435)	(0.00100)
L3 Stringency Index	0.00147***	0.00484***	0.00103***	0.000571*	0.000712
	(0.000496)	(0.000758)	(0.000384)	(0.000308)	(0.000703)
L1 Economic Support Index	0.00292***	-0.00198***	-0.000666**	-0.000432	0.000454
	(0.000425)	(0.000651)	(0.000330)	(0.000266)	(0.000609)
L2 Economic Support Index	0.000394*	-0.00299***	-0.00227***	0.000181	0.000163
	(0.000239)	(0.000371)	(0.000183)	(0.000155)	(0.000351)
L3 Economic Support Index	-0.000774**	0.00215***	0.00152***	-0.00000258	-0.0000194
	(0.000353)	(0.000558)	(0.000265)	(0.000229)	(0.000523)
Formality F.E	No	No	Yes	Yes	Yes
Occupation F.E	No	No	Yes	Yes	Yes
Firm Size F.E	No	No	Yes	Yes	Yes
Individual Characteristics F.E	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes

Notes: table shows regression results estimated from equations 4.12 and 4.13. Occupation (ISCO-08), formality status, firm size, and individual characteristics fixed effects are included. Individual fixed effects include race, age, sex and education of the individual. In the first column (for the estimates on employment), I only include individuals with a legal working age, i.e. those at least 14 years old. Controls include concurrent and lagged (up to three quarters) of COVID-19 cases and deaths related to COVID-19, and lagged levels (up to three quarters) of the Stringency Index and Economic Support Index. Robust standard errors in parentheses , *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Heterogeneous Impact of Government Measures By Individual Characteristics As a second step, I aim to uncover the uneven effects of stringency and economic support measures based on individual characteristics. To do so, I estimate equation 4.14, where in addition to the variables as explained before, X_i are individual level

characteristics.[30]

$$Y_{i,T} = \nu + \beta_2 X_i \times C_{UF,T} + \sum_{h=0}^{3} \beta_h C_{UF,T-h} + \sum_{h=0}^{3} \phi_h Cases_{UF,T-h} + \sum_{h=0}^{3} \eta_h Deaths_{UF,T-h} \tag{4.14}$$

$$+ \omega_1 Y_{i,T-1} + \epsilon_{i,T}.$$

I report the coefficients of β_2 in table I.8. I find that for every one unit increase in the economic support index or Stringency Index, informal workers were more negatively impacted by the lockdown measures than their formal workers counterparts, perhaps due to the more contact intensive occupations for informal workers. This is evident from the lower hours worked and labour income earned. For every one unit increase in the Economic Support Index or Stringency Index, secondary and tertiary educated workers have higher wages than primary educated workers. The negative impact on female's hours worked and monthly wage due to the implementation of the economic support measures and the stringency measures was higher than those for males. Also, for a one unit increase in either the Economic Support Index or the Stringency Index, younger cohorts of workers were more likely to work more hours than their older counterparts, but received lower wages. Further, workers at small firms were more likely to be negatively impacted by increases in either index than workers in larger firms. For a unit increase of the economic support index and Stringency Index, black or brown individuals earned less labour income than their white counterparts. Workers in teleworkable occupations saw an increase in hours worked and income earned with increases in either index relative to workers in less teleworkable sectors. This finding can be reflecting that with higher levels of lockdown measures for example, workers in more teleworkable sectors saw an improvement in their incomes and an increase in the amount of hours worked. However, this increase could be reflecting those more able workers who managed to remain in employment. The evidence presented here corroborates the findings in the previous sections. The impact of the pandemic and the associated policies to curb the spread of the virus has been highly uneven, affecting females, younger cohorts, informal workers and small firms more negatively than others.

[30]I also focus the analysis on the third quarter of 2021 for the following reasons: 1- This is the most recent available quarter of data, with the least number of recorded COVID cases and allows me to assess the lingering impact of containment measures, and 2- I can control for lagged income and hours worked per individual to attenuate some of the bias of estimating a highly endogenous relationship.

Table I.8: Heterogeneity of the Impact of Government Response on Hours Worked and Labour Income

Economic Support Index	(1) Log Diff in Hours	(2) Log Diff in Income
C. X Informal	-0.000254	-0.00000259
	(0.000248)	(0.000551)
C. X Female	0.0000784	-0.000615
	(0.000198)	(0.000451)
C. X age 14-23	0.000566**	0.0000278
	(0.000275)	(0.000627)
C. X age 24-34	0.000458*	-0.00123**
	(0.000270)	(0.000609)
C. X ≤10 employees	-0.000484*	0.00121**
	(0.000275)	(0.000602)
C. X >10 and ≤50 employees	-0.000445	0.00147*
	(0.000346)	(0.000756)
C. X Teleworkability	0.000415*	-0.000605
	(0.000236)	(0.000524)
C. X Contact Intensity	0.00000525	0.00000295
	(0.00000564)	(0.0000127)
Stringency Index	0.000632	-0.000168
	(0.000396)	(0.000901)
L1. X Informal	-0.000316	0.000287
	(0.000260)	(0.000579)
L1. X Female	-0.000182	0.000541
	(0.000206)	(0.000477)
L1. X age 14-23	0.0000248	0.000516
	(0.000291)	(0.000668)
L1. X age 24-34	-0.0000967	0.00146**
	(0.000286)	(0.000649)
L1. X ≤10 employees	0.000141	-0.00122*
	(0.000293)	(0.000642)
L1. X >10 and ≤50 employees	0.000274	-0.00137*
	(0.000370)	(0.000811)
L1. X Teleworkability	-0.000211	0.000413
	(0.000252)	(0.000562)
L1. X Contact Intensity	-0.00000642	-0.00000431
	(0.00000642)	(0.0000145)
Race and Education F.E	Yes	Yes
Occupation F.E	Yes	Yes
Controls	Yes	Yes

Stringency Index	(1) Log Diff in Hours	(2) Log Diff in Income
C. X Informal	0.0000368	0.0000839
	(0.000472)	(0.00105)
C. X Female	-0.000130	0.000682
	(0.000393)	(0.000895)
C. X age 14-23	0.000481	0.000743
	(0.000541)	(0.00122)
C. X age 24-34	0.000937*	0.00120
	(0.000530)	(0.00119)
C. X ≤10 employees	-0.000655	-0.00144
	(0.000540)	(0.00118)
C. X >10 and ≤50 employees	0.000183	0.000746
	(0.000672)	(0.00146)
C. X Teleworkability	0.00112**	0.000646
	(0.000467)	(0.00103)
C. X Contact Intensity	0.00000231	-0.0000112
	(0.0000104)	(0.0000232)
Economic Support Index	0.000372	-0.000557
	(0.000320)	(0.000725)
L1. X Informal	-0.000347	0.000114
	(0.000340)	(0.000754)
L1. X Female	0.0000567	-0.000519
	(0.000285)	(0.000648)
L1. X age 14-23	-0.0000210	-0.000223
	(0.000390)	(0.000881)
L1. X age 24-34	-0.000469	-0.000759
	(0.000383)	(0.000859)
L1. X ≤10 employees	0.000238	0.00108
	(0.000391)	(0.000855)
L1. X >10 and ≤50 employees	-0.000229	-0.000471
	(0.000487)	(0.00106)
L1. X Teleworkability	-0.000679**	-0.000492
	(0.000339)	(0.000748)
L1. X Contact Intensity	-0.00000302	0.0000104
	(0.00000768)	(0.0000171)
Race and Education F.E	Yes	Yes
Occupation F.E	Yes	Yes
Controls	Yes	Yes

Notes: The tables show regression results based on the specification in equation (4.14) and PNAD-Continua microdata. Each table represents a separate estimation, one with the Economic Support Index interacted with individual characteristics and another with the Stringency Index interacted with individual characteristics. Standard errors are in parenthesis. Survey data design and sampling weights are taken into consideration in the estimation. Controls include contemporaneous and lagged values of number of COVID cases and COVID related deaths, as well as lagged values of the government measure under study (whether the Stringency Index or the Economic Support Index). The omitted categories are as follows: on education it is tertiary education, on age it is 54+ cohort, on race it is white, and on firm size it > 30 employees. Robust standard errors in parentheses , *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Impact of Government Measures on Sectoral Employment Finally, I estimate the impact of the Stringency Index and Economic Support Index on sectoral employment using equation (4.15). I look at 15 aggregated sectoral levels exploiting the variation in employment and containment measures across seven quarters and 27 Federation Units.[31] I use the notation that $\Delta Y_{s,t,UF}$ represents the change in sectoral employment between quarter-year t, for sector s and Federation Unit UF. Taking the difference is important because it accounts for pre-COVID employment levels. My measure of policy is $c_{t,UF}$, which corresponds either to the Stringency Index or the Economic Support Index. In this analysis, μ_{UF} are Federation Unit fixed effects, and γ_t are year-quarter Fixed Effects. To attenuate reverse causality and omitted variable bias, I control for lagged values of the containment measures (both of them) $c_{t-1,UF}$ where $c \in \{\text{Stringency}_{t-1,UF}, \text{Economic Support}_{t-1,UF}\}$, contemporaneous and lagged COVID cases and related deaths and a range of time varying UF characteristics ($X_{t,UF}$). $\theta_{1,s}$ captures the point estimate of interest that represents the contemporaneous impact of a specific containment measures.

$$\Delta Y_{s,t,UF} = \mu_{UF} + \gamma_t + \theta_{1,s} c_{t,UF} + \theta_{2,s} c_{t-1,UF} + \sum_{h=1}^{2} \phi Cases_{t-h,UF} + \sum_{h=1}^{2} \phi Deaths_{t-h,UF} \tag{4.15}$$

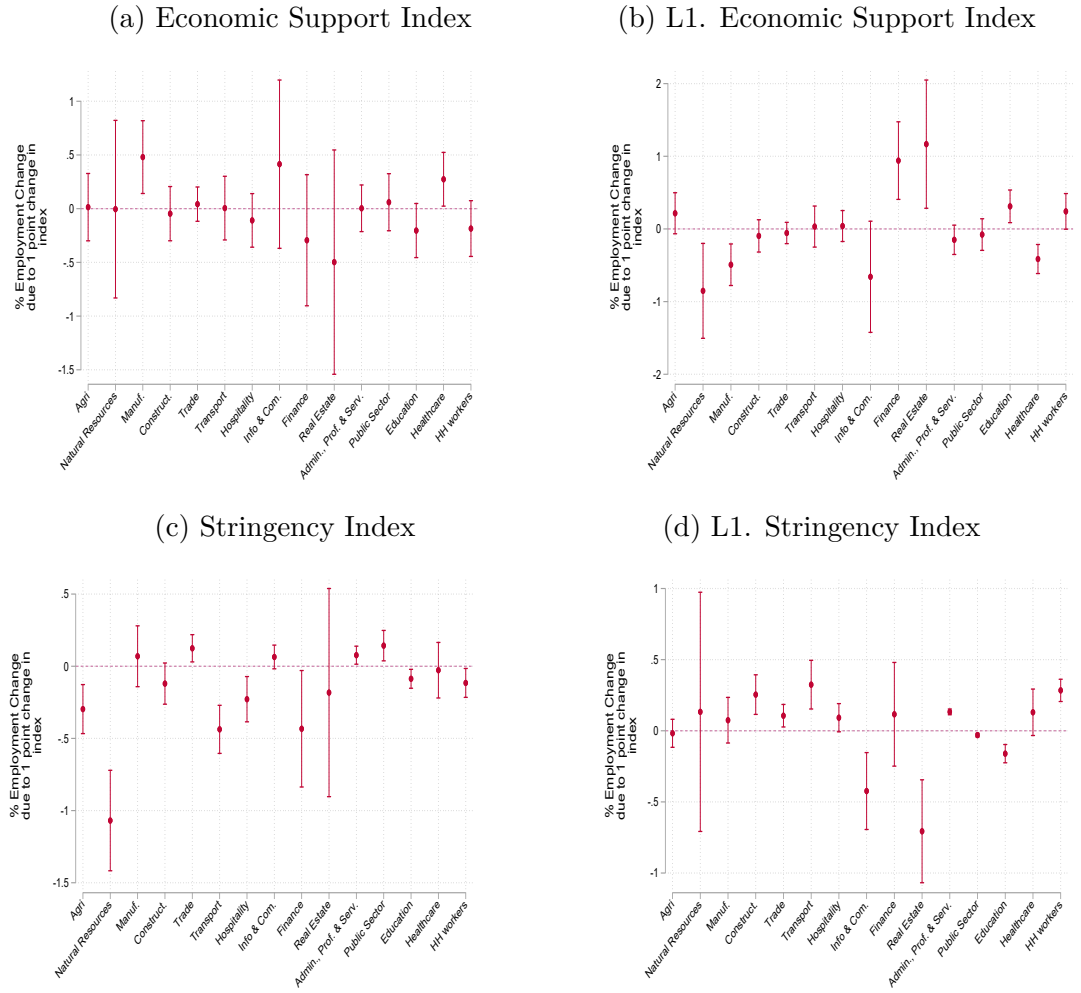
$$+ \tau X_{t,UF} + \epsilon_{s,t,UF}.$$

I report the point estimates of $\theta_{1,s}$ and the 90% confidence intervals in figure I.36. I find that the Stringency Index impacted negatively employment in the transport, education and household workers sectors. These results are expected, considering the impact that the stringency measures have on mobility (Transport) and contact-intensive sectors (Household workers). The impact on education is interesting and could be attributable to the inability of certain schools, perhaps those located in rural areas, to shift and maintain remote teaching. Further, the sign of the point estimates (negative) on the employment in hospitality and trade is consistent with expectation, albeit insignificant. The Economic Support Index significantly and adversely impacted workers in the public, education and household sectors but had a positive impact on

[31] This is a short time dimension and a limitation for this analysis. Ideally this analysis is done monthly, but I do not have sectoral monthly employment data. Also ideally, the analysis can be done at the municipal level but I also lack the data at this level for the different sectors.

the manufacturing sector. The positive impact on the manufacturing sector could be due to the formal worker retention scheme captured in the Economic Support Index.

Figure I.22: Percent Change in Employment by Sector in Response to Government Measures



Notes: The graphs show the point estimates and confidence intervals of $\theta_{1,s}$ and $\theta_{2,s}$ in equation (4.15) for each government response measure (Stringency Index, and Economic Support Index). The equation is estimated for each one of the sectors on the x-axis. Source: PNAD-Continua and author's estimations. 90 percent confidence intervals are depicted around the point estimates.

4.2 Auxílio Emergencial

Having shown the impacts of the aggregate government response on labour market outcomes and sectoral employment across Federation Units, this section focuses on one of the main economic support measures. This social support program was "Auxílio Emergencial", the generous cash transfer program offered for eligible informal and

unemployed individuals. Auxílio Emergencial (EA) was one of the most generous, robust and extensive social protection programs implemented across developing countries (Lara de Arruda et al., 2021).

This section is divided into two subsections. The first explains Auxílio Emergencial and highlights the innovations adopted in this program. It is worthwhile to shed light on this because of the importance of this cash transfer in alleviating the impact of the pandemic, and serves as a learning experience in how developing countries can design and implement social transfer programs when faced with shocks with such reverberating consequences. The second and third subsections estimate the distributional impacts of the emergency aid as well as some key counterfactual scenarios. In the final section of this paper, I study this program in more depth and go a step further to estimate its distributional impacts and effects on poverty and inequality.[32]

Throughout the period that it was introduced, Auxílio Emergencial reached approximately 44.7% of the population and approximately 40% of households.[33] The program aimed to reduce the impact of the pandemic on the most vulnerable and, in addition, to reduce the prevalence of poverty and inequality (as measured by the Gini index).[34] The program was robust, and it built on the infrastructure of existing social programs to help identify and deliver payments, which permitted its fast roll-out across the country. Eligibility for the EA was automatically evaluated for those individuals who were already in the Unified Civil Registry (Cadastro Único, CU). This substantially simplified the implementation of the EA since the CU already contained data on a large fraction of the poorest Brazilians, including all Bolsa Família beneficiaries and those who had applied to the program but were not beneficiaries yet. This feature of Auxílio Emergencial served to lower the risk of errors of exclusion (Lara de Arruda et al., 2021). The benefit from Auxílio Emergencial does not allow the recipient to have other forms of social transfers, except for Bolsa Família. Its main focus is on extending support to those who are chronically and temporarily poor and not covered by social

[32]As mentioned earlier, another important program that was implemented at that time is the "Emergency Employment and Income Maintenance Benefit -BEM" that was targeted at formal workers. BEM is a financial benefit offered to workers in the formal sector who faced a temporary (90-120 days) reduction in hours worked or income earned or suspension in employment caused by the pandemic. The financial benefit is provided by the Federal Government (CAIXA, 2021). It would have been ideal to include a detailed analysis of this program similar to the one done of "Auxílio Emergencial". Unfortunately, I do not have data on this program, but it is a fruitful avenue for future work when the data becomes available.

[33]My measure of individual coverage consists of those people who are direct beneficiaries plus those who belong to a family where at least one member of the household received the aid.

[34]The conclusions drawn from reductions in inequality are to be interpreted carefully because I am relying on household self-reported surveys for my estimation. Thus, the reported income earned may be subject to under-reporting bias by the wealthier population (Souza, 2016; Picketty, 2013).

insurance system nor by labour-related guarantees in the country. The emergency aid is given at the individual level based on individual characteristics, however, a cap exists on the total number of family members that can receive the aid. This created a duality between coverage rates expressed according to different units of analysis (household or individual).

4.2.1 Description of Auxílio Emergencial

Auxílio Emergencial was enabled due to a major law that allowed more flexibility in Brazilian fiscal rules to deal with the COVID-19 pandemic. The law concerning Auxílio Emergencial was approved within 6 days with an extraordinary budget of BRL 541.1 billion (USD 103.46 billion).[35]

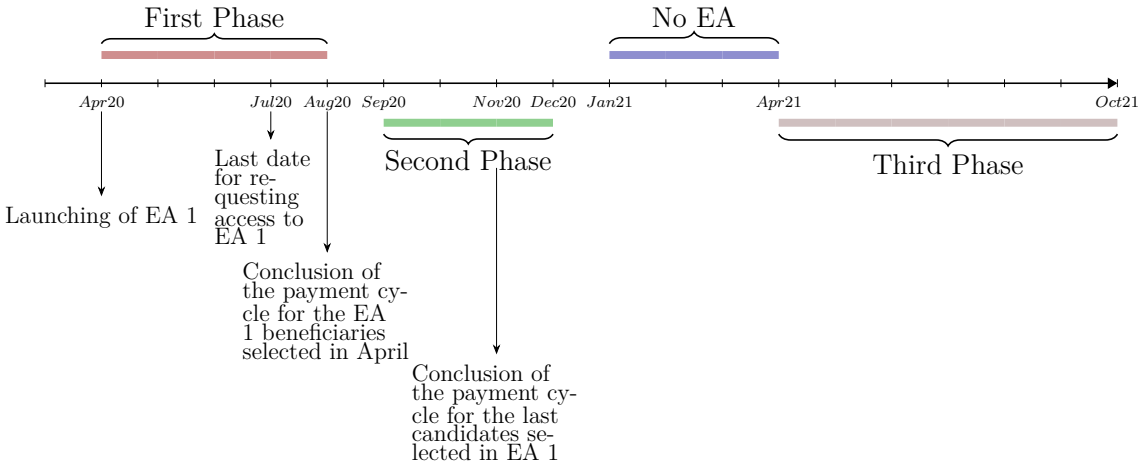
With the expectation that the pandemic impacts will be short lived, the emergency aid was established for a limited temporal horizon. In its first phase, Auxílio Emergencial was meant to last for three months only. However, it ended up being extended across three phases (EA1, EA2, and EA3). The eligibility criteria as well as the amount provided varied as the phases of the program progressed. The first stage (April to July 2020) was designed to last for three months and guaranteed a payment of $R 600 per month for up to two individuals in a family, and this amount was doubled for female heads of single parents with children. To be eligible for the programme, recipients must not be formally employed, are unemployed, or are an individual micro-entrepreneur and have a monthly per capita family income of up to half the minimum wage ($R 522.5 or 99.9 USD). The second stage (August to November 2020) reduced the benefits by half and left the eligibility criteria the same. The third phase of the emergency cash transfer from April to July 2021 restricted the payment benefits to only one person per household with a further reduced amount of R$ 250, with additional $R 125 for female head of households with children. Table I.18, in the appendix, provides a detailed description of the eligibility and amounts per phase of the program. What is most notable throughout these phases is that the special focus on female headed households receiving double the amount of money remained in place. Further the eligibility protocols became more robust as the program advanced, reaching approximately 40 major administrative registries in the cross-verification process.

The program weighed heavily on the fiscal balance sheets with expenditure to combat the pandemic reaching a soaring $R 541.1 billion (cite the period)), with 63% of

[35]Social assistance represents 63 percent of these extraordinary expenses, of which 87 percent was allocated to the EA.

this budget going to social assistance, of which 87% was allocated to Auxílio Emergencial. The three phases of the program cost as follows: $R 68.3 million for the first phase, $R 56.8 million for the second and $R 39.4 million for the third. To put the generosity of the program in perspective, the Auxílio Emergencial amount was approximately three times that of Bolsa Familia and equivalent to two thirds the average monthly income of Bolsa Familia beneficiaries for most recent years. This amount is also equal to 50% of the national minimum wage and approximately equal to 50% of the average amount received by beneficiaries eligible for the Unemployment Insurance, Seguro Desemprego. The program reached all 5,570 municipalities in Brazil, but the coverage was uneven across them. Table I.19 in the appendix shows the distribution of recipients for Auxílio Emergencial across the 27 Federation Units. The data show that 7.28% of the aid beneficiaries reside in São Paulo, which is relatively low, considering that São Paulo alone has more than 21% of the total population in Brazil. Only 20.27 % of EA recipients lived in rural areas.

Table I.9: Timeline of the three Auxílio Emergencial (EA) Phases



Notes: The figures shows a timeline of the different phases of Auxílio Emergencial since its establishment until October 2021. Source: Information is based on Arruda et al. 2021.

4.2.2 Distributional Impacts and the Labour Market Effects of Auxílio Emergencial

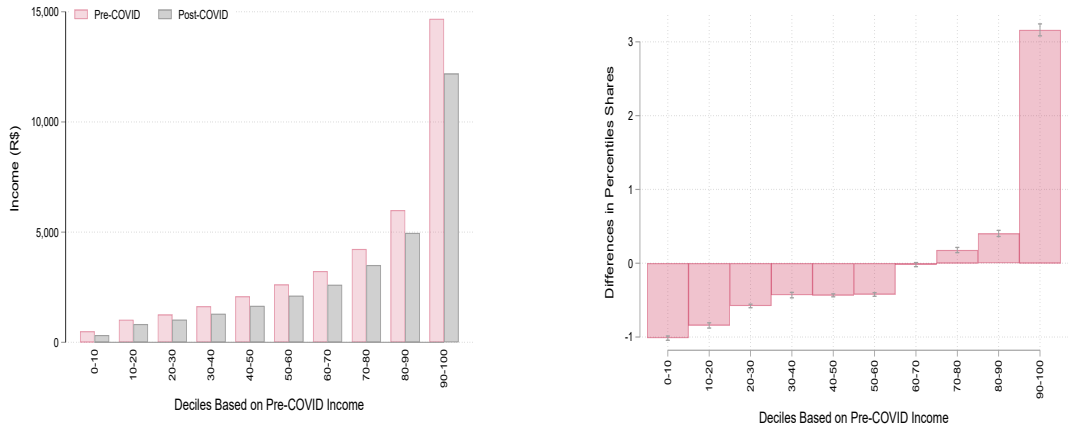
Having described the importance, eligibility, and phases of the implementation of Auxílio Emergencial, I discuss in this section its actual distributional consequences

on household incomes. I then make use of the differential amounts given to female single mothers versus their male counterparts to provide evidence on the labour market consequences of this large income increase through a gender lens. I conclude the analysis on Auxílio Emergencial in section 4.2.3, where I study the impact of the program on inequality and poverty, and present a counterfactual analysis on the role of better targetting. I have shown in section 3.3 evidence that economically more vulnerable groups (informal workers, women, and younger cohorts of workers) were the hardest hit by the COVID-19 shock on Brazil's labour market. As shown in figure I.23, while labour income fell across the entire income distribution, poor households were the most affected. Households in the bottom income decile saw their monthly labour income decrease by 35 percent from R$502 to R$327, resulting in a loss of about 1 percentage point of their percentile share in total income. Labour income in the top decile of the distribution decreased by 17 percent, resulting in an increase in their percentile share by approximately 3.2 percentage points.

Figure I.23: Distribution Analysis based on post-COVID Labour Income

(a) By Income Decile

(b) Change in Percentile Shares vs Pre-COVID



Notes: figure I.23(a) Pre-COVID income is the baseline. Figure depicts the average labour income pre and post COVID by decile. Pre-COVID income is the baseline. Figure I.23(b) depicts the differences in percentile shares based on labour income between pre and post COVID by decile. Source: PNAD COVID (May 2020).

Distributional Impacts of Auxílio Emergencial As mentioned earlier, in order to mitigate the impact of the pandemic on vulnerable households, the government rolled out the EA cash transfer program in April. The EA initially intended to cover 67.7 million eligible beneficiaries, around one-third of the population.

Table I.10 shows the characteristics of people who belonged to a household that received the aid, disaggregated by formality status, race and education. The aid beneficiaries were 92.4% informal workers, consistent with the targeting of the program; of those, 46.7% were women. Some 55% of beneficiaries are brown, and the majority of beneficiaries have not completed schooling beyond high school. The labour income wage for those beneficiaries remains very close or below to the equivalent of the minimum wage with 38 hours work week. The program seems to have been well targeted, and in what follows I investigate the impact per income deciles.

Table I.10: Baseline Characteristics of Individuals Receiving Auxílio Emergencial

	Mean			Percent		
	Income Pre Covid	Hours Pre Covid	Age	Female (%)	Removed Quarantine (%)	Remote Work (%)
Color or race						
White	2,900.9	40.1	35.9	22.93	3.08	21.15
Black	1,803.0	39.5	36.2	4.33	0.82	4.43
Asian	3,307.8	40.4	39.9	0.41	0.08	0.39
Brown	1,799.0	39.1	32.8	23.26	3.69	22.74
Indigenous	1,728.1	38.8	34.1	0.16	0.03	0.16
Ignored	1,745.9	41.7	32.8	0.02	0.00	0.02
Education						
Without instruction	937.2	36.2	19.8	5.29	0.10	5.37
Incomplete elementary school	1,260.0	39.0	33.4	14.74	1.38	15.52
Complete Elementary	1,578.6	40.1	47.2	3.37	0.58	3.20
Incomplete high school	1,399.7	39.1	28.6	4.57	0.74	4.96
Complete high school	1,737.5	40.4	38.0	12.45	2.79	11.59
Incomplete higher	2,134.0	38.7	30.5	3.02	0.62	2.76
Graduated	3,932.4	39.4	42.1	6.15	1.23	4.43
Postgraduate, Master's or Doctorate	6,666.1	39.8	43.9	1.52	0.26	1.06

Notes: This table depicts the baseline characteristics of beneficiaries of the Auxílio Emergencial cash transfer by income, hours worked age, sex, removal from job and remote work. Analysis is dis-aggregated by formality status, race and education. Formality status is based on whether or not the worker holds a formal contract. With the PNAD COVID survey, this is the only available indicator on the formality status. Source: PNAD COVID (May 2020).

Analysis based on PNAD-COVID for May (Figure I.24) shows that more than 70 percent of households in the bottom decile received the aid, accounting for almost 23 percent of total EA disbursements.[36] In total, the bottom three deciles received 53.9 percent of benefits. However, there was room for improvement in targeting, as the top three deciles of the income distribution received about 10.7 percent of the aid. As

[36]Deciles are calculated on equal number of households.

of May, at the peak of the crisis, the transfers increased the average income of the bottom 40 percent of households by 20 percent relative to the level of pre-COVID income (Figure I.25).

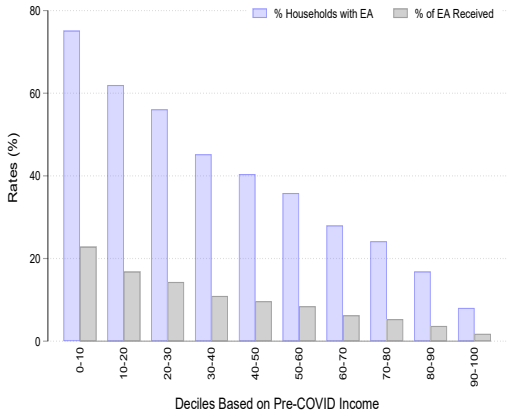
The aid corresponded to 150 percent of the pre-COVID labour income in the bottom decile and 40 percent of the pre-COVID overall median labour income. The EA more than compensated for the negative impact of the pandemic on poverty and income, at least temporarily (Figure I.25). Inequality as measured by the Gini index in May would have increased from 0.53 pre-COVID to 0.58 post-COVID. Likewise, the poverty headcount ratio based on the extreme poverty line of R128 per month (USD 1.9 per day) would have increased sharply from 4.7 percent to 13.9 percent in May—corresponding to approximately 15.7 million individuals saved from falling into extreme poverty.[37] Instead, with the EA, the poverty headcount ratio fell to 4.4 percent and the Gini coefficient to 0.51—both lower than their pre-COVID levels (Figure I.26(b)). For households headed by single-earner women with at least one child, who received double the transfer, the poverty rate would have increased from 11 to 30.4 percent without EA; instead it temporarily fell to 8.4 percent because of the EA.[38] Depending on the poverty line used, EA prevented between 7.8 and 16.4 million people from falling below the poverty line between pre-COVID and post-COVID income from May until October.

The counterfactuals without EA could admittedly overestimate poverty and income inequality outcomes as the initial amount of the benefit could have discouraged labour participation for workers in the bottom deciles (who, as discussed, received on average 150 percent of their pre-COVID income), particularly women head-of-households (who received double the benefit). An analysis per month throughout the first and second phases of Auxílio Emergencial shows that there is already evidence of a gradual rebound in the labour markets shown by the progressive decrease of the post-COVID poverty headcount ratios between May and October, along with a slight increase in the ratio including EA in October that reflects the lower benefit granted in September. This is consistent with the recovery of monthly wages and hours worked for both males and females over the same period.

[37] Based on households that have at least one income earning member. Pre-COVID income differs from month to month due to small errors of reporting of this value. So, in figure I.26(b), I take the arithmetic mean across all the months.

[38] This poverty rate is calculated as the total of members of households headed by a female who is a single earner and who are under the extreme poverty line, divided by the total of such households in Brazil.

Figure I.24: Share of Total Households and Share of Total EA Paid per Decile

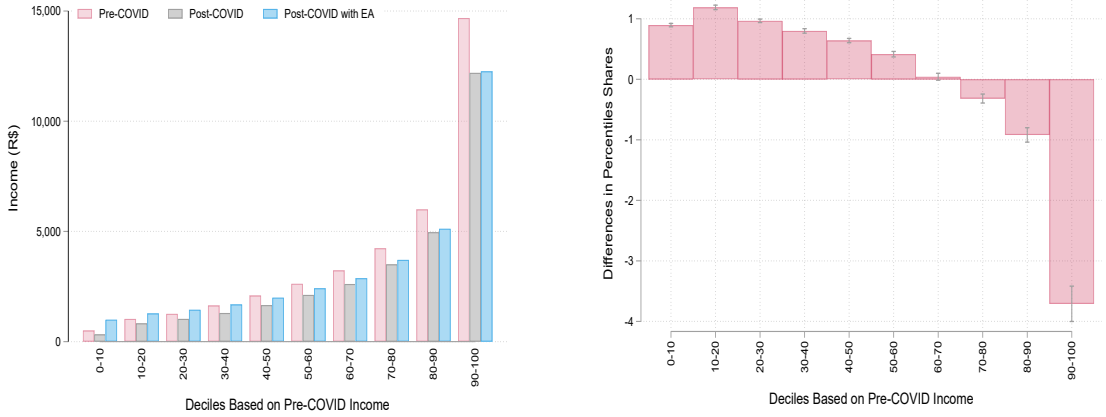


Notes: The figure depicts the percent of households that received the Auxílio Emergencial (EA) and the percent of Auxílio Emergencial (EA) received, by income deciles. The calculation of the deciles is based on Pre-COVID income. Source: PNAD COVID (May 2020).

Figure I.25: Income Distribution with EA by Income Decile

(a) Pre-COVID and Post-COVID with and Without EA Labour Income

(b) Change in Percentile Shares vs Pre-COVID



Notes: figure I.25(a) depicts the average labour income pre, post COVID and post COVID with Auxílio Emergencial transfer by decile. The calculation of the deciles is based on Pre-COVID income. Figure I.25(b) depicts the differences in percentile shares based on labour income between pre and post COVID with EA transfers by decile. The calculation of the deciles is based on Pre-COVID income. Source: PNAD COVID (May 2020).

Gendered Labour Market Impacts of Auxílio Emergencial Having shown that Auxílio Emergencial represented a significant positive income shock to households at the poorest deciles, I ask how this transfer could have impacted differently the labour market outcomes for male- versus female-headed households. This is especially interest-

ing because, in the first phase of the program, female-headed households who are single mothers received twice the benefit of their male-headed households counterparts.[39]

I focus on labour market outcomes in the last month available to me in the PNAD COVID survey, November 2020, which also marks the end of the second phase of Auxílio Emergencial. The analysis presented herein is descriptive in nature.[40] I conduct the analysis in two steps. In the first step, I compare the labour outcomes of female heads of households against the labour outcomes of those male heads of households that received the aid. I also compare both against female heads of households that didn't receive the aid. In the second step, conditional on having received the aid in May, I compare the labour outcomes in November of female single parents with male single parents. Note that female heads of households, who are also single parents, received benefits twice as large as the male-headed households. A limitation of this latter comparison is that male single parents are likely to be a heavily selected subset of all adult males.

I estimate this analysis at the individual level, controlling for income earned in May and pre-COVID income. In order to do so, I make use of the nature of the survey to match individuals from two months: May to November, i.e. from the first months EA was given, until the last month of the second phase of EA.[41] I estimate the following equations and report the results in table I.11:

[39]Those are all male headed households, and not necessarily those who are single parents or have a household of comparable size.

[40]It goes beyond the scope of this paper to estimate the causal impact of Auxílio Emergencial on labour market outcomes, and this is a future avenue of research. That said, I have thought and attempted to implement two identification strategies to estimate a causal impact. The first of those is a regression discontinuity design that exploits the income eligibility cut-off. However, I find that there is bunching around the threshold which indicates that the self-reported income around the threshold might have been manipulated. The second strategy is a difference in difference approach, however, I cannot exploit a staggered rollout as Brazil implemented the program nationally at the same time.

[41]This survey uses outgoing rotation groups from the regular PNAD Continua (ILO Policy Brief, 2021), and the sample is fixed for all months until the end of the survey (Pesquisa Nacional por Amostra de Domicílios PNAD COVID19, 2020). I note that not all individuals match across the two months of May and November despite merging at a very strict level of Federation unit, day, month, and year of birth, sex, race, household interview number, primary sampling unit, position in the household and urban versus rural regions. Therefore it is important to conduct a balance test to understand if the matched sample is representative of the entire sample of households in Brazil. I report the results of this balance test in table I.22. Based on normalised differences as recommended by Imbens and Wooldridge (2009), I find that the matched and the unmatched samples are similar on observable characteristics.

$$(Y_{i,Nov}|EA_{May}=1) = \mu_1 + Female_{Head} \times EA_{Nov} + income_{i,pre-COVID} + income_{i,May} + hours_{i,pre-COVID} + hours_{i,May} + \lambda_{education} + \lambda_{race} + \lambda_{Area} + \lambda_{UF} + \epsilon_{i,Nov}. \tag{4.16}$$

$$(Y_{i,Nov}|EA_{May}=1) = \mu_1 + Female_{SingleParent} + EA_{Nov} + income_{i,pre-COVID} + income_{i,May} + hours_{i,pre-COVID} + hours_{i,May} + \lambda_{education} + \lambda_{race} + \lambda_{Area} + \lambda_{UF} + \epsilon_{i,Nov}. \tag{4.17}$$

In these equations, $Y_{i,T}$ represents the labour outcomes in terms of income, informality status, hours worked, worked remotely, as well as health outcomes such as COVID-19 symptoms and hospitalisation due to COVID-19, in November. First, I compare female versus male head of households that received the aid in the first month of the program, in May. I find that, relative to men, female heads of households that received the aid in November earned less income in November, and worked less than their male counterparts. Further, female heads of households who also received the aid in November, were more likely to be employed informally, less likely to do remote work, and more likely to exhibit COVID-19 symptoms and to be hospitalised as a result.

Next, I turn my attention to single heads of households and single parents. For females within this classification and who satisfy the eligibility requirements of EA, the benefit amount received was double those of males. Therefore it is an interesting comparison to ask whether, conditional on having received the aid in May, females' labour market outcomes differ from those of males. I find that, despite having received double the amount of aid as males in May, female single heads of households, who are also single parents, remained at a disadvantage in the labour market relative to male single parents. Those female heads of household earned less income, worked fewer hours, and were less likely to work remotely, and more likely to be informally employed. Health related outcomes also show that those females exhibited more COVID-19 symptoms

and were more likely to be hospitalised as result.[42] Since in November 2020, the pandemic in Brazil was still unfolding and the COVID-19 cases increasing, it is a short timeline to see the effects of this differential impact between males and females of this one-time increase in income through COVID-19 on labour market outcomes and health outcomes. A promising avenue in this research is to attempt to relate individuals who received EA with future labour market outcomes in 2021, after the recovery of the market.[43]

4.2.3 Poverty, Inequality and Counterfactual Analysis

In this section I turn my attention to study the impact of Auxílio Emergencial on headcount poverty and inequality. As noted earlier, Auxílio Emergencial was well targeted, but leakage did occur, and some richer households received the benefit. The program also came at a very large fiscal cost (4% of GDP). Therefore, I present counterfactual implementation scenarios at different levels of the fiscal commitment and targeting options, to see the resulting changes in the poverty headcount ratio.

Poverty and Inequality Under Auxílio Emergencial As Brazil is classified as a middle income country (Lara de Arruda et al., 2021), I calculate the poverty headcount ratio based on the international upper middle income poverty line (5.5$ 2011 PPP) and report the results in figure I.26(a). In order to put the analysis in historical perspective, I also estimate the headcount poverty rate for the quarterly data using PNAD-Continua.[44] I also estimate, using the same poverty line, the female poverty headcount ratio for phase one and two of EA, and I simulated phase three of EA.

As can be seen in figure I.26(a), the poverty headcount ratio estimated increased drastically from 24.77% in the third quarter of 2019 to 30.36% in May 2020. However, with the implementation of Auxílio Emergencial, the sharp increase in poverty headcount ratio returned to a level below the pre-pandemic levels. When other social

[42] I find similar results when estimating the impact on labour market outcomes of the other the members of the households and not just the heads of households.

[43] This can be challenging due to the rotational nature of the PNAD Continua surveys and the fact that data on EA comes from PNAD COVID which was discontinued in November, 2020. An attempt to merge households based on a created unique individual identifier that includes the primary sampling unit, Federation Unit, Household selection number, position in the household, day, month and year of birth, sex, and race results in merging less than 10% of the sample between PNAD COVID November, 2020 and the 1st quarter of PNAD Continua, 2021.

[44] This is with the caveat that the two datasets are not directly comparable due to their method of collection and time coverage. I am reassured that my calculations on headcount poverty rate match those calculated by Brazil's official statistics agency *O Instituto Brasileiro de Geografia e Estatistica* (IBGE).

Table I.11: Impact on labour Income and Hours Worked

	(1) Log(Income)	(2) Informal Employee	(3) Hours Worked	(4) Remote Work	(5) COVID-19 Symptoms	(6) Hospital if COVID-19
Female vs Male Heads						
Female Head X EA in Nov	-0.243***	0.0955***	-2.988***	-0.0913***	0.0729***	0.0131***
	(0.0194)	(0.00563)	(0.230)	(0.0227)	(0.00550)	(0.00242)
Age	0.00149***	-0.00245***	-0.0766***	0.00224***	-0.000221	-0.0000166
	(0.000565)	(0.000165)	(0.00665)	(0.000617)	(0.000161)	(0.0000708)
log(Income Pre-COVID)	0.578***	-0.0405***	1.666***	-0.0707***	0.000330	0.00127
	(0.0107)	(0.00304)	(0.125)	(0.0113)	(0.00297)	(0.00130)
log(Income in May)	0.0716***	0.00641***	-0.234***	-0.0101***	-0.0000701	0.00118***
	(0.00293)	(0.000814)	(0.0344)	(0.00353)	(0.000795)	(0.000350)
Hours Worked Pre-COVID	0.000903	-0.00182***	0.376***	0.00248***	0.000996***	0.000416***
	(0.000622)	(0.000180)	(0.00733)	(0.000723)	(0.000176)	(0.0000774)
Hours Worked in May	0.00252***	0.000112	0.231***	0.000664	-0.000770***	-0.000421***
	(0.000394)	(0.000119)	(0.00465)	(0.000521)	(0.000116)	(0.0000510)

	(1) Log(Income)	(2) Informal Employee	(3) Hours Worked	(4) Remote Work	(5) COVID-19 Symptoms	(6) Hospital if COVID-19
Female vs Male Single Heads and Single Parents						
Female Single Head and Parent X EA in Nov	-0.204***	0.143***	-4.536***	-0.165***	0.0678***	0.00815*
	(0.0341)	(0.00925)	(0.396)	(0.0378)	(0.00963)	(0.00436)
Received EA in November	-0.141***	0.0205**	-0.232	0.00120	0.00528	-0.00406
	(0.0293)	(0.00836)	(0.339)	(0.0312)	(0.00870)	(0.00394)
Age	0.000778	-0.00263***	-0.0897***	0.000598	-0.00110***	0.0000274
	(0.00120)	(0.000329)	(0.0138)	(0.00129)	(0.000343)	(0.000155)
log(Income Pre-COVID)	0.579***	-0.0357***	1.511***	-0.0691***	0.00560	0.00608**
	(0.0199)	(0.00541)	(0.231)	(0.0205)	(0.00563)	(0.00255)
log(Income in May)	0.0923***	0.00544***	-0.266***	-0.00465	0.000396	0.00106
	(0.00556)	(0.00149)	(0.0645)	(0.00692)	(0.00156)	(0.000704)
Hours Worked Pre-COVID	0.00192	-0.000554*	0.359***	-0.000327	0.000908***	0.000385**
	(0.00118)	(0.000320)	(0.0136)	(0.00136)	(0.000334)	(0.000151)
Hours Worked in May	0.000907	0.000234	0.226***	0.00216**	-0.000512**	-0.000700***
	(0.000741)	(0.000209)	(0.00859)	(0.000960)	(0.000218)	(0.0000984)
Education F.E	Yes	Yes	Yes	Yes	Yes	Yes
Race F.E	Yes	Yes	Yes	Yes	Yes	Yes
Area F.E	Yes	Yes	Yes	Yes	Yes	Yes
UF F.E	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The table shows regression results estimated from equation 4.16 and 4.17. Education, race, area and Federation Unit (UF) fixed effects are included in the estimations. Informal employee, remote work and hospital if COVID-19 are binary variables, were 1 indicates that the individual was informally employed, worked remotely or was hospitalised due to COVID-19. Robust standard errors in parentheses , *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

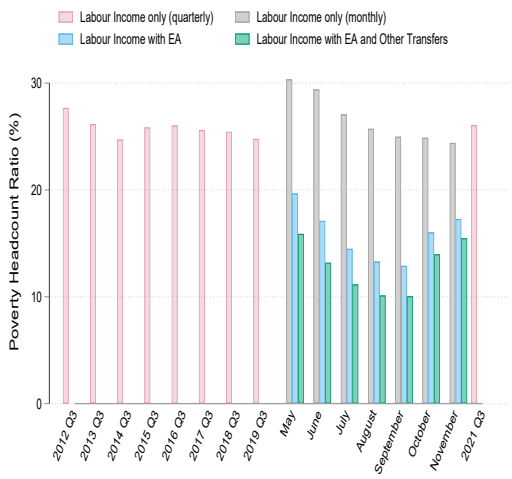
transfers are included in addition to labour income in the calculation of the welfare measure, the poverty headcount ratio fell to as low as 10.05% in September. The adjusted Gini coefficient decreased from 0.54 in the third quarter of 2019 to 0.528 in 2020, perhaps due to the implementation of the generous Auxílio Emergencial. In the third quarter of 2021, which is the most recent data available, the poverty headcount ratio rose again to 26.06%, a level higher than the pre-pandemic levels since the third quarter of 2013. With fiscal sustainability in mind, I also entertain a counterfactual scenario (not shown) where the amount of aid is cut in half for each month from May to November. The poverty headcount ratio and inequality as measured by the Gini coefficient would have fallen below pre-pandemic levels, with the former reaching at low as 15.85% and the latter 0.488 in September, a reduction that is also supported by a recovery in labour incomes across the months studied (May to November).

Further, the consumer price index in Brazil increased by 12.5% by the third quarter of 2021 and by 2.63% in 2020, compared to the same quarter in 2019. I calculate the increase in the poverty headcount ratio due to inflation and I find that in 2021, the poverty headcount ratio would have risen due to inflation by approximately 2.93 percentage points which added approximately 6.2 million people to the poor, and 0.98 percentage points in 2020 with added 2.07 million to the poor, had inflation stayed the same as it was in the third quarter of 2019.

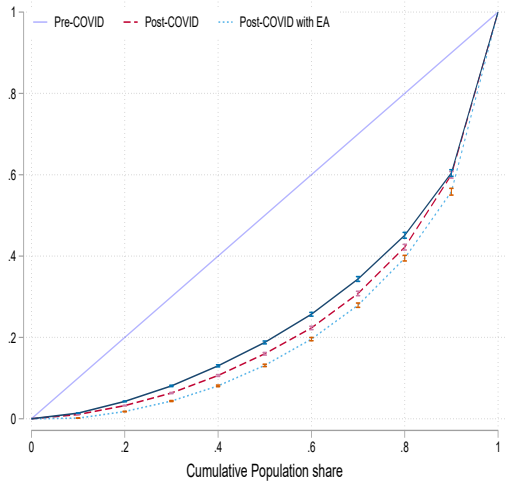
Moreover, Auxílio Emergencial in its first and second phases provided double the amount of the benefit to female single heads of households with children compared to baseline benefit given to other eligible households. With the implementation of this transfer and the special focus on female head of households, the female headcount poverty ratio was reduced on average across all months by 13.84 percentage points compared to what would have happened had the Auxílio Emergencial not been implemented. The female poverty headcount ratio was also 9.52 percentage points less than that pre-Covid on average across all months. I estimate that with the reduction of this benefit in phase three of the program as well as the abolition of EA and the introduction of Auxílio Brazil, poverty among females would have risen on average across the months studied by 8 percentage points in comparison with the headcount poverty ratio achieved with Auxílio Emergencial (figure I.27).

Figure I.26: Poverty Headcount Ratio (%) and Inequality

(a) Poverty Headcount Ratio (%)

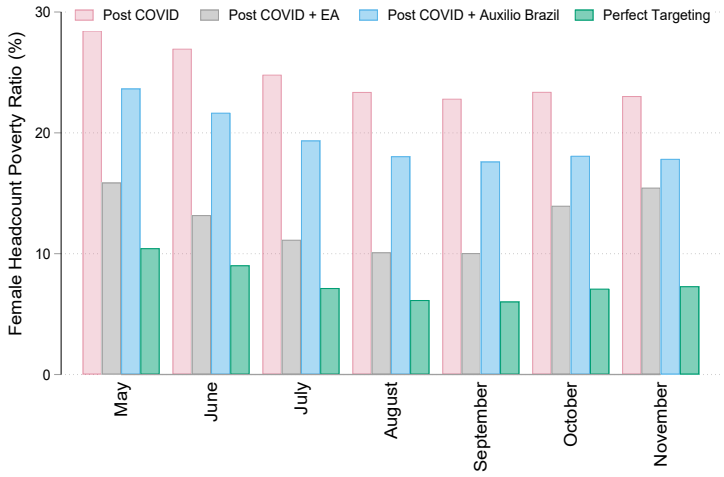


(b) Income Inequality (Gini Index)



Notes: Figure I.26(a) depicts the evolution of poverty headcount ratio. Estimates are based on international upper middle income poverty lines (5.5$ 2011 PPP), also used by IBGE. Pre-COVID poverty headcount ratio is based on reported labour income in PNAD-Continua. Other social transfers included in the calculation of poverty headcount ratio from May to November include, in addition to Auxílio Emergencial, Bolsa Familia, income from donations, alimony, income from retirement, unemployment insurance and other like rent. Source: PNAD COVID and PNAD-Continua. Figure I.26(b) depicts the Gini index based on labour income, pre, post COVID with and without EA transfers. Source: PNAD COVID (May 2020).

Figure I.27: Female Poverty Headcount ratio Counterfactual Analysis for the Administration of EA



Notes: The figure depicts the evolution of female poverty headcount ratio. Estimates are based on international upper middle income poverty lines (5.5$ 2011 PPP), also used by IBGE. The ratio is calculated for the different phases of the program and simulated for the third phase of the program (EA 3) where data is unavailable. Female poverty headcount ratio is calculated as the ratio of all members of female headed households that fall below the poverty line as a fraction of all members of female headed households.

Counterfactual Scenarios of the Implementation of Auxílio Emergencial

Given the difficulty of ensuring that a program of the magnitude of Brazil's EA is paid only to those who meet the criteria, while at the same time all eligible people receive the benefit - thus minimising errors of inclusions and exclusions respectively - I analyse below different targeting scenarios for the month of October and their impact on poverty.

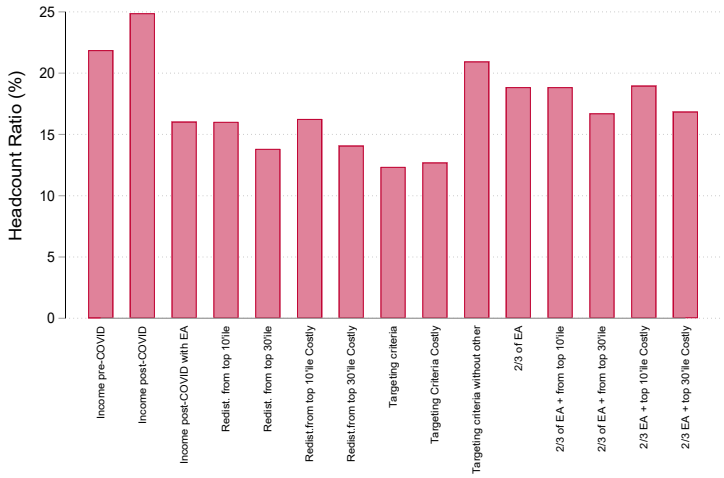
The first three columns of figureI.28 show the headcount ratio (extreme poverty line) for households earning labour income pre-COVID, post-COVID, and post COVID with the EA. As noted above, the EA was key in preventing a rapid and significant rise in short-term poverty levels, and infact lowering the poverty headcount compared to pre-COVID times. I simulate the headcount ratio that would have prevailed if all payments that were received by the top 30th percentile were redistributed to the bottom decile without any costs, and with an administrative targeting cost of 3 percent of the aid given, respectively (bars 4,5). In this latter simulation, the cost of the targeting is passed down to the beneficiary in the form of less aid received. As expected, in both cases, the extreme poverty headcount ratio is reduced to below its observed level, to about 1.2 percent.

In bars 6, 7, and 8, I replicate the eligibility criteria for the EA targeting and present results for a counterfactual scenario where only those that are eligible as per the stipulated conditions receive the aid with no extra cost of targeting, and with a 3 percent cost of targeting. Under the former simulation, the total amount spent would have been R$3.1 billion lower than what was disbursed in October. In these scenarios of the targeting criteria simulations, the headcount ratio falls to 0.8 and 0.9 percent respectively—still less than the current headcount ratio with EA of 1.9 percent.

I also present the scenario of perfect targeting but with incorporating an additional requirement stating that those who receive other welfare benefits except for Bolsa Familia are not eligible for the EA aid. I do not observe that this criterion is binding in practice, whereby 4.2 million households said they received EA alongside benefits such as unemployment benefits and retirement pension. In this scenario, the poverty headcount ratio would rise to 4.7 percent, if individuals receiving benefits other than Bolsa Familia were excluded.

Finally, I simulate counterfactual scenarios where the unit transfer is reduced by a third (to R$400). Holding the original distribution of transfers, would result in R4.9 billion savings in October (approximately 33 percent of the EA cost in the same month) and a marginal increase in the extreme poverty headcount ratio to 2.7 percent as transfers are reduced. With redistribution from the top 30th percentile to the bottom deciles, and with a 3 percent targeting cost, the extreme poverty ratio would be reduced to 1.9 and 2 percent of the population, respectively. Hence with better targeting, even two thirds of the original amount of aid would be very effective in reducing extreme poverty to a ratio very close to the one observed in October.

Figure I.28: Poverty Counterfactual Analysis for the Administration of EA



Notes: The figure depicts the headcount poverty ratio at the national equivalent of the international upper middle income poverty lines (5.5$ 2011 PPP) for different counterfactual scenarios. Source: Author's analysis of PNAD COVID for the month of October.

5 Conclusion

This paper provided an extensive and detailed analysis of the short-term and medium-term impacts of the COVID-19 pandemic on labour markets in Brazil. To conduct this analysis, I use a new monthly household survey dataset, PNAD COVID, as well as a quarterly household survey, PNAD Continua. The paper is divided into two parts. In the first part, I assess the impact of the pandemic on labour market characteristics. I present an ex-ante analysis of jobs and wages at risk using a framework by Alfaro et al. (2020) that studies jointly the demand and supply shocks, focusing on the role of informal workers and small firms, both of which are key characteristics of the Brazilian labour market. I compare the results of this modelling exercise with an ex-post analysis that disentangles the margins of heterogeneity and looks at realised impacts across gender, education, occupation, urban or rural regions and formality status. The second part of this paper assesses the impact of Brazil's policy responses, considering both lockdown measures and economic support. A number of government programs had significant impacts on the recovery of income and hours worked – although these impacts were highly differentiated across the economy. In the latter section, I delve deeper into the generous emergency cash support programme implemented by the government and I study its impact on poverty and inequality. I take a step further and

discuss what would have happened to poverty under different counterfactual scenarios regarding the implementation of this emergency program.

My analysis shows that close to 25 million jobs were at risk in the early stages of the pandemic, considering inter-sectoral linkages and the effect of both the supply and demand channels. The employment-at-risk was mainly concentrated in sectors that are less teleworkable and more contact-intensive such as trade, transport and hospitality. Informal workers face almost double the employment risk than formal workers because they work in more contact-intensive sectors that are more subject to lockdowns. A comparison with actual changes in employment shows that, while the sectoral distribution of employment-at-risk is fairly well estimated, total job losses were significantly lower than the universe of employment-at-risk, in particular for small and medium-sized firms. While several factors are at play, the government's employment retention program, which covered around 10 million formal workers and allowed for a subsidised reduction in hours or temporary suspension of the employment relationship, as well as the loan programs dedicated for the support of small and medium enterprises, likely played an important role. I find that the reduction in employment disproportionately affected informal workers, women and young cohorts of workers. Older cohorts of workers retained their employment but were more likely to be temporarily removed from their jobs, with high reductions in hours worked and monthly wages. In the 3rd quarter of 2021, labour markets have seen some recovery, especially for informal workers, but the adverse impact on females' wages and hours worked persisted, indicating a scarring effect.

While labour income fell across the entire income distribution, poor households were the most affected. Households in the bottom income decile lost about 1 percentage point of their percentile share in total income as their monthly labour income decreased by 30 percent. My analysis shows that poverty headcount ratio and the Gini index would have increased sharply as a consequence, to about 30 percent and 0.58 respectively, in May, compared to about 22 and 0.53 pre-COVID. However, the government sponsored Auxílio Emergencial program was successful in cushioning the blow of the crisis on the livelihoods of Brazilians, and reduced headcount poverty ratio and income inequality to 16% and 0.51 respectively, lower than their pre-COVID levels. The played an important role in meeting an urgent need to protect the livelihoods of those most severely impacted, the informal workers, and female head of households.

Nevertheless, at a cost of about 4 percent of GDP, the Auxílio Emergencial program was never intended to be a long-term program and arguably is fiscally unsustainable

beyond the short-term. The program was terminated in October 2021. Given the significant role played by the EA in mitigating the short-term increase in poverty and income inequality, a substantial improvement in the labour market will be necessary to avoid a potentially sharp deterioration in social outcomes in the longer term. Encouragingly, the reduction by half of the EA transfer resulted in only a marginal up-tick of the poverty ratio in October. To gauge options to provide continued stimulus, I present counterfactual scenarios for the distribution of the EA under different targeting and benefit options. I find that further reducing the transfer by one-third (to R400 a month per individual) with improved targeting would be successful in keeping the poverty headcount ratio close to its October levels even after accounting for the administrative cost of targeting.

Going forward, given the importance of the federal expenditure ceiling to preserve debt sustainability, macroeconomic stability and market confidence, it will be important that any additional social assistance spending be financed within the expenditure limits imposed by the ceiling. Considering the rigidity of Brazil's budget, this will require firm implementation of fiscal structural reforms aimed at providing flexibility for the budget process and reducing mandatory spending to create the necessary fiscal space for increased social assistance spending.

Acknowledgements I am very grateful to Fredrick Toscani, Valentina Flamini and Carlo Pizzinelli for their excellent collaboration, support and guidance. I am also thankful to Nicolas Landeta and Nischel Pedapudi for excellent administrative support, and to IMF colleagues for insightful comments and suggestions.

Appendices of Chapter 1

I.A Supplementary Material: Ex-Ante Accounting Analysis

This section presents the calibration used in the ex-ante accounting analysis in section 3.2. $Lock_s$ reflects the impact of different stages of lockdowns in each sector. The following values are assumed following the calibration by Komatsuzaki et al. (2020):

Table I.12: The calibration measures of $Lock_s$ and $Lock_{s_{selective}}$

Sector Code	Sector Name	$Lock_s$	$Lock_{s_{selective}}$
A	Agriculture	0	0
B, D, E	Mining, electricity and water	1	1
C	Manufacturing	0.681	0
F	Construction	1	0
G	Wholesale and Retail	0.561	0.280
H	Transportation and Storage	0.647	0.647
I	Accommodation and Food Services	1	1
J	Information and Communication	0	0
K	Finance and Insurance	0	0
L	Real Estate	0	0
M, N, R, S	Professional services, administrative services, arts and entertainment, etc.	0.114	0.114
O	Public administration	0	0
P	Education	0	0
Q	Human Health and Social Services	0	0
T	Household as employer etc.	1	0.7

Notes: The values are used from Alfaro et al. 2020 (p.33) classification of lockdown probabilities across sectors, and map the sectoral classification of Colombia in the original paper (the GEIH data) to that of Brazil and aggregate to single digit sectors as shown in the table (Komatsuzaki et al., 2020).

S_{it} takes the following values:

Table I.13: S_{it} Values Across the Shutdown Stage and the Two Opening Stages

	Shutdown period	Partial Sectoral Opening Period		Advanced Sectoral Opening Period	
		Opening<50%	Opening>50%	Opening<50%	Opening>50%
Small, self-employed, informal	0.154	0.231	0	0	0
Medium	0.104	0.156	0.052	0.26	0
Large	0.047	0.0705	0.0235	0.011	0

Notes: The exact values for the shutdown period are approximated from the answers to the BICS survey "'How much confidence does your business have that it will survive the next three months?". I combine the probability, by firm size, for those that answered that they have "No confidence" or "Low Confidence" (Lambert and Van Reenen, 2021).

The direct demand shock $Direct_{Loss_s}$ is as adapted from Alfaro et al. (2020) classification and follows the calibration by Komatsuzaki et al. (2020).

Table I.14: $Direct_{Loss_s}$ (Direct Demand Shock)

A	Agriculture	0.055
B, D, E	Mining, electricity and water	0
C	Manufacturing	0.111
F	Construction	0
G	Wholesale and Retail	0.188
H	Transportation and Storage	0.526
I	Accommodation and Food Services	0.540
J	Information and Communication	0
K	Finance and Insurance	0
L	Real Estate	0
M, N, R, S	Professional services, administrative services, arts and entertainment, etc.	0.081
O	Public administration	0
P	Education	0
Q	Human Health and Social Services	0
T	Household as employer etc.	0

Notes: This table adopts the consumption shock analogous to that of Sweden to Brazil's sectoral classification. AD=1.692, following Alfaro et al. (2020) and Komatsuzaki et al. (2020). Sources: Input-Output table: sourced from WIOD for Brazil.

Distribution of employees by contact intensive jobs and teleworkable jobs, subdivided by gender and education.

Table I.15: Distribution of Workers by Sex and Education and Occupation Teleworkability and Contact Intensity

Sex	Contact Intensity	Total Employed	% Share of Employment	Teleworkability	Total Employed	% Share of Employment
Male	Low	33123165	35	Low	39314250	42
Female	Low	25591454	27	Low	26953671	29
Male	High	19082368	20	High	12891283	14
Female	High	15537081	17	High	14174864	15

Education Level	Contact Intensity	Total Employed	% Share of Employment	Teleworkability	Total Employed	% Share of Employment
Primary	Low	20508285	22	Low	28250746	30
Secondary	Low	22324122	24	Low	29066176	31
Tertiary	Low	15882213	17	Low	8950999	10
Primary	High	10500374	11	High	2757913	3
Secondary	High	14956804	16	High	8214749	9
Tertiary	High	9162270	10	High	16093485	17

Notes: This table shows the distribution of employees by low and higher contact intensive and teleworkable jobs, subdivided by gender and education. The table shows the total number of employees as well as the share of total employment in (%). Low Teleworkability and low Contact Intensity characterizes all occupations with an index on each < 0.5. Source: PNAD-Continua and Dingel and Neiman (2020).

The following table shows the characteristics and job attributes for formal and informal workers. I consider seven characteristics: distribution over firm size, gender, Education, distribution over contact intensive and teleworkable occupations, urban or rural areas, and percentage of employees with income above or below the minimum wage.

Table I.16: Characteristics and Job Attributes for Formal and Informal Workers

	Formality Status		
	Formal	**Informal**	**Total**
	%	%	%
Firm Size			
1 to 5 people	15.4	83.5	50.8
6 to 10 people	13.2	6.6	9.7
11 to 50 people	20.7	4.8	12.4
51 or more people	50.7	5.1	27.0
Gender			
Male	56.1	57.8	57.0
Female	43.9	42.2	43.0
Education			
No education or less than 1 year of study	0.8	3.2	2.0
Incomplete elementary	13.4	31.3	22.3
Complete elementary	6.6	9.6	8.1
Incomplete high school	5.8	8.0	6.9
Complete high school	38.9	29.1	34.0
Incomplete higher education	7.4	5.6	6.5
Graduated	27.2	13.3	20.3
Contact Intensive Occupation			
Low Contact Intensity	65.0	60.4	62.7
High Contact Intensity	35.0	39.6	37.3
Ability to Telework			
Low Teleworkability	64.9	77.5	71.2
High Teleworkability	35.1	22.5	28.8
Urban/Rural			
Urban	94.1	82.3	88.2
Rural	5.9	17.7	11.8
Labour Income to Minimum Wage			
Above	96.8	59.5	78.2
Below	3.2	40.5	21.8

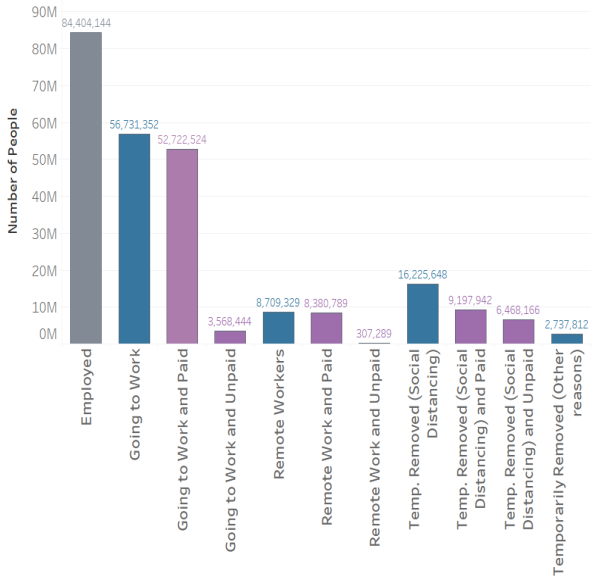
Note: This table shows the percentage of workers within each firm size, by sex, by education level and type of occupation (ability to telework or contact intensive) and monthly income relative to minimum wage. Informal workers are those that do not have a formal contract (are either hired informally by formal firms or informal self employed), are not registered with the "National Register of Legal Entitites", and do not contribute to the mandatory social security payment on the job. Source: PNAD-Continua.

I.B Supplementary Material: Employment Statistics During the Pandemic

In this section, I present further descriptive analysis on employment and labour market outcomes during the pandemic.

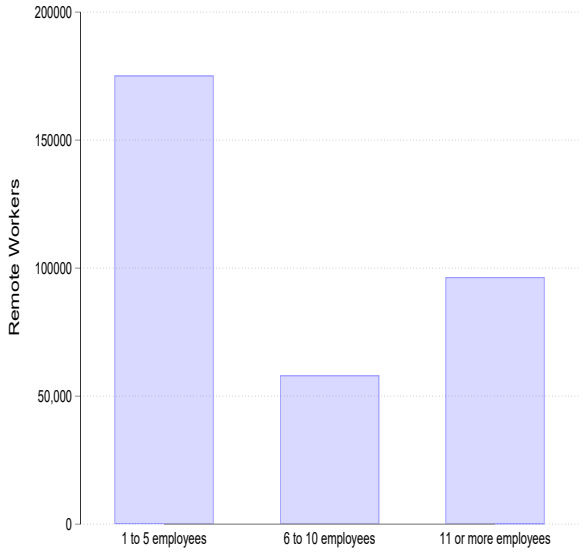
I present the decomposition of employed people across different categories: whether going to work, or working remotely among others. I show that at the onset of the pandemic there was 8.7 million remote workers, which is approximately 10% of total employment. I also show that there was approximately 16.2 million workers who were temporarily removed due to social distancing.

Figure I.29: Employment Decomposition in May



Notes: figure depicts the decomposition of employed people in May across different dimensions. Source: PNAD COVID (May, 2020).

Figure I.30: Remote Employment by Firm Size in May

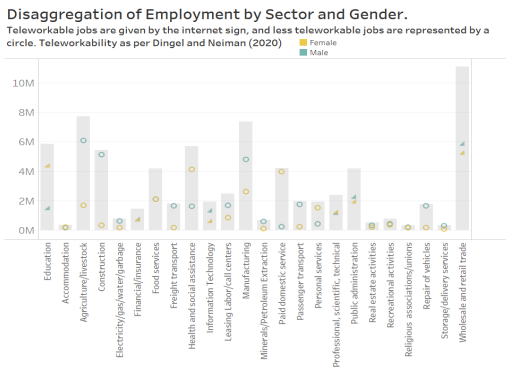


Notes: figure depicts the number of people employed remotely by firm size. A limitation in this analysis is that there exists a lot of missing values on the firm size for employees, therefore these statistics are representative for workers who answered this question. Source: PNAD COVID (May, 2020).

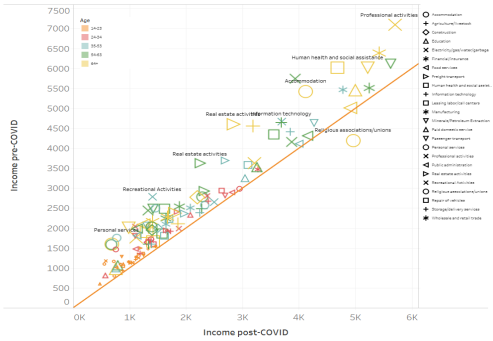
I use PNAD COVID for May in order to decompose the employment of individuals across sectors, and divided by gender (Figure I.31(a)). I also depict graphically, for different age groups and across occupations, the average income pre-COVID and post-COVID (Figure I.31(b)).

Figure I.31: Distribution of Workers across Sectors with Margins of Heterogeneity

(a) Dis-aggregation of Employment by Sector and Gender



(b) Income pre-COVID versus Income post-COVID by sector and age



Note: figure I.31(a) The numbers shown in this figure are calculated from PNAD COVID May 2020. The three bars indicate that the sector is more teleworkable and the circle indicates that the sector is more contact intensive. Figure I.31(b) shows numbers that are calculated from PNAD COVID May 2020.

I show the change in labour outcomes between pre and post COVID for employed people. I consider this change for change in weekly hours, monthly wage and hourly wage for total employed people, formal employees, and for males and females.

Table I.17: Change in Income and Hours Worked

	(1) All Employed	(2) Employed (excluding zero wage earners and zero hours worked)
Total		
Monthly Income Change (%)	-17.9	-10.8
Change in Hours (%)	-30.3	-11.2
Hourly Income Change (%)	17.9	0.4
Formal Employees		
Change in Hours-Formal Employees (%)	-25.9	-7.9
Change in Income-Formal Employees (%)	-6.4	-4.3
Hourly Income Change-Formal Employees (%)	26.3	3.9
Males and Females		
Change in Hours - Male (%)	-26.6	-10.4
Change in Hours - Female (%)	-35.9	-12.5
Change in Income - Male (%)	-17.9	-11.2
Hourly Income Change-Male (%)	11.8	-0.9
Change in Income - Female (%)	-17.7	-10.1
Hourly Income Change-Female (%)	28.3	2.7

Note: The table presents the percent change in hours worked per month, wage per month and wage per hour for the month May (comparing pre-COVID and Post-COVID). The first column presents the percent changes among all employed people and the second column presents the results for employed people excluding those that report earning a zero wage and working zero hours.

I.C Supplementary Material: Auxílio Emergencial

I.C.1 Details on the Auxílio Emergencial Cash Transfer Programme

EA was implemented by the Ministry of Citizenship (Ministério da cidaddnia), while the Social Security Technology and Information Company (Dataprev) was responsible for the EA eligibility criteria verification. By May 2020, the government had analysed all 52 million individuals in the CU; of these, 29.7 million were approved, including 96 percent of Bolsa Família beneficiaries, who were switched to the EA. The rest of the beneficiaries were identified by self-selection through an app run by Caixa Economica Federal. This is the largest state-owned bank in Brazil (by number of customers) and also responsible for the payment of Bolsa Família benefits. Beneficiaries also had the option of receiving the cash transfer via a digital savings account, which also had the benefit of promoting fintech and financial inclusion during the early stages of the pandemic. Further, in order to improve selection and eligibility procedures, EA formed a "Management Commission", composed of high-level members of the Ministry of Citizenship and participant auditing agencies, such as the Tribunal de Contas da União (Federal Court of Auditors, TCU) and the Controladoria Geral da União (General Controller's Office, CGU) of the Federal Government (Lara de Arruda et al., 2021).

Admission into Auxílio Emergencial was done across three channels. First, there was an automatic selection of Bolsa Familia beneficiaries who would be eligible to receive the EA transfer. Second, there was an automatic selection of individuals registered in the largest administrative registry of social programs (The single registry or Cadastro Ùnico or CadÙnico) who were not Bolsa Famalia beneficiaries. The administration utilised a static extract of this registry dating to April 2, 2020. This means that there is potential for exclusion risk for individuals and families whose income or family composition changed since then. And third, there was passive selection based on submitted applications to the mobile app developed by CAIXA for this purpose.[45] This app would allow individuals whose situation was not accurately reflected in the Single Registry to apply for the transfer; however, they have to declare the same family

[45]This led to the creation of a new temporary registry called Pùblico Beneficiàrio do Auxìlio Emergencial fora do Cadastro Ùnico (Public not Registered within the Single Registry, ExtraCad).

composition as listed in the Single Registry. This means that if the total number of beneficiaries within a family reached the maximum specified in each phase of the program, as indicated in the Single Registry at that point in time, then new applications from new members of this family will be rejected (Lara de Arruda et al., 2021).

Table I.18: Eligibility and Benefits of Auxílio Emergencial

EA wave	Eligibility	Benefit	Total Budget spent
EA 1: May to Aug 2020	2 beneficiaries per household Double Benefit for Single Mother Not formally employed, unemployed, MEI. Income $\leq$ R$ 522,50 per family member or total family income $\leq$ R$ 3,135 Have a valid CPF number; Not receive any other Government aid except Bolsa Família.	BRL 600 (USD 114.72) .	R$ 231 Billion
EA 2: Sep to Nov 2020	Similar to EA 1 except individual residing abroad and imprisoned no longer eligible.	BRL 300 (4 $\times$ USD 57.36) EA 2 reduced by 50 % and discontinued the double amount payments for single mothers, and living in families with two beneficiaries.	R$ 64 Billion
EA 3: Apr to Oct 2021	Similar to EA 1 and 2. Added: should not be a trainee, medic, or beneficiary of study grants.	EA 3 restricted to one person per family and reduced the benefit amount to BRL 250 (USD 47.80), with additional BRL 125 (USD 23.90) for single mothers, and a further reduction of BRL 100 (USD 19.12) for single-member families with no children.	R$ 43 Billion

Notes: The table describes for each phase of Auxílio Emergencial, the eligibility criteria, the benefits of the social transfer program and the total budget spent on each of the three phases. Single mothers refers to a single female head of household who is also a single mother. Any other government aid includes, for example, retirement, pension, unemployment insurance, disability benefits, Continued Payment Benefits. MEI refers to individual micro entrepreneurs, and CPF refers Individual Taxpayer's Registry. Source: Information is based on Lara de Arruda et al., 2021.

I.C.2 Descriptive Evidence on Auxílio Emergencial

This section presents further descriptive evidence on the social support program, Auxílio Emergencial.

In figure I.32, I show the percent of households that received aid at each level of the aid. The non-zero percentage of households (approximately 3%) that received an amount of aid equalling 1800$R is surprising and is indicative of leakage because the maximum amount that can be received is 600 $R for a maximum of two members in the household.

Table I.19: Distribution of Auxílio Emergencial across Federation Units as a % of Total Population

	Emergency aid related to coronavirus		
	Yes	**No**	**Total**
	% of Total	% of Total	% of Total
Federation unity			
Rondônia	0.42	0.43	0.85
Acre	0.24	0.18	0.42
Amazonas	1.20	0.72	1.91
Roraima	0.14	0.12	0.26
Pará	2.65	1.45	4.10
Amapá	0.27	0.13	0.40
Tocantins	0.43	0.32	0.75
Maranhão	2.28	1.08	3.36
Piauí	0.98	0.58	1.56
Ceará	2.62	1.73	4.35
Rio Grande do Norte	0.98	0.69	1.67
Paraíba	1.14	0.77	1.90
Pernambuco	2.60	1.93	4.53
Alagoas	1.01	0.57	1.59
Sergipe	0.65	0.45	1.10
Bahia	4.32	2.74	7.07
Minas Gerais	4.14	5.94	10.08
Espírito Santo	0.83	1.09	1.92
Rio de Janeiro	2.97	5.25	8.23
São Paulo	7.28	14.64	21.92
Paraná	1.95	3.50	5.45
Santa Catarina	0.84	2.59	3.43
Rio Grande do Sul	1.57	3.84	5.40
Mato Grosso do Sul	0.52	0.77	1.30
Mato Grosso	0.71	0.94	1.64
Goiás	1.49	1.88	3.37
Distrito Federal	0.47	0.98	1.45

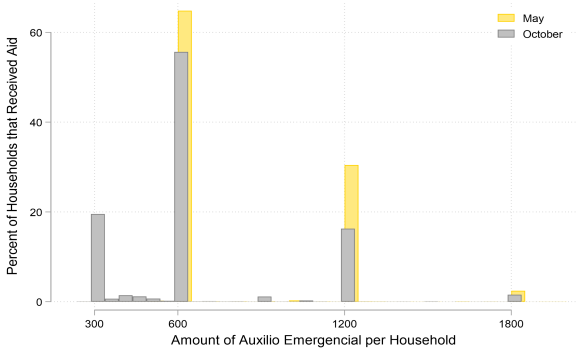
Notes: This table depicts the % of Auxílio Emergencial beneficiaries per every Federation Unit as a proportion of the total population in Brazil. Source: PNAD COVID (May 2020).

Table I.20: Distribution of Auxílio Emergencial across Federation Units as a % of UF Population

	Emergency aid related to coronavirus		
	Yes	**No**	**Total**
	% of pop	% of pop	% of pop
Federation unity			
Rondônia	49	51	100
Acre	58	42	100
Amazonas	63	37	100
Roraima	54	46	100
Pará	65	35	100
Amapá	67	33	100
Tocantins	57	43	100
Maranhão	68	32	100
Piauí	63	37	100
Ceará	60	40	100
Rio Grande do Norte	59	41	100
Paraíba	60	40	100
Pernambuco	57	43	100
Alagoas	64	36	100
Sergipe	59	41	100
Bahia	61	39	100
Minas Gerais	41	59	100
Espírito Santo	43	57	100
Rio de Janeiro	36	64	100
São Paulo	33	67	100
Paraná	36	64	100
Santa Catarina	25	75	100
Rio Grande do Sul	29	71	100
Mato Grosso do Sul	40	60	100
Mato Grosso	43	57	100
Goiás	44	56	100
Distrito Federal	32	68	100

Notes: This table depicts the percentage of Auxílio Emergencial beneficiaries per every Federation Unit as a proportion of the total population per Federation Unit. Source: PNAD COVID (May 2020).

Figure I.32: Percentage of Households by Aid Value



Notes: figure depicts the percentage of households receiving aid at every level of aid. Source: PNAD COVID.

In what follows, I show further evidence of the presence of leakage in the Auxílio Emergencial social support program for households that do not satisfy the income eligibility criteria.

Table I.21: Aid Received by Top 2 Deciles Households

	Mean	Median	Min	Max
Auxilio Emergencial	659	600	300	6000
Household Members	2.6	2	1	12
Pre-COVID Income	5475.6	4600	100	52200
Post-COVID Income	5162.6	4300	1250	42000
Post-COVID Income with EA	5821.6	5100	1580	42600

Source: PNAD COVID and authors' calculations

In section 4.2, I merge individuals between the two datasets of PNAD COVID for the months of May and November. I merge along the following variables: day, month and year of birth, gender, race, the primary sampling unit, role in the household, household survey selection number, Federation Unit and urban or rural regions. However, not all the individuals in May merge to those in November. Since my analysis is done on the merged sample only, it is important to understand how this matched sample differs from the unmatched sample on observable characteristics to be able to make a statement on whether the analysis is representative of all individuals. I consider four observable characteristics: age, gender, race, and education. I report the t-test p-value as well as the normalised difference. I find that, based on the normalised differences,

there is no statistically significant difference across the matched and unmatched sample on observable characteristics.[46]

[46]The normalised differences measure is preferred to the T-test P-value because the latter can be highly sensitive to the sample size. I follow Imbens and Wooldridge (2009) and consider significant differences as those with a normalised difference exceeding 0.25.

Table I.22: Balance Tests for PNAD COVID

	(1)		(2)		T-test	Normalized
	Unmatched Sample		Matched Sample		P-value	difference
Variable	N	Mean/SE	N	Mean/SE	(1)-(2)	(1)-(2)
Male	256885	0.486	195329	0.497	0.000***	-0.022
		(0.001)		(0.001)		
Female	256885	0.514	195329	0.503	0.000***	0.022
		(0.001)		(0.001)		
White	256885	0.402	195329	0.470	0.000***	-0.139
		(0.001)		(0.001)		
Black	256885	0.093	195329	0.084	0.000***	0.033
		(0.001)		(0.001)		
Asian	256885	0.008	195329	0.007	0.000***	0.016
		(0.000)		(0.000)		
Brown	256885	0.492	195329	0.436	0.000***	0.112
		(0.001)		(0.001)		
Indigenous	256885	0.004	195329	0.002	0.000***	0.025
		(0.000)		(0.000)		
Not Specified	256885	0.001	195329	0.000	0.000***	0.033
		(0.000)		(0.000)		
Without instruction	186109	0.127	195329	0.092	0.000***	0.113
		(0.001)		(0.001)		
Incomplete elementary school	186109	0.343	195329	0.277	0.000***	0.140
		(0.001)		(0.001)		
Complete Elementary	186109	0.069	195329	0.060	0.000***	0.035
		(0.001)		(0.001)		
Incomplete high school	186109	0.097	195329	0.094	0.000***	0.012
		(0.001)		(0.001)		
Complete high school	186109	0.221	195329	0.251	0.000***	-0.072
		(0.001)		(0.001)		
Incomplete higher	186109	0.048	195329	0.067	0.000***	-0.084
		(0.000)		(0.001)		
Graduated	186109	0.078	195329	0.123	0.000***	-0.152
		(0.001)		(0.001)		
Postgraduate, Master's or Doctorate	186109	0.017	195329	0.037	0.000***	-0.125
		(0.000)		(0.000)		
Resident's age	256885	34.368	195329	33.058	0.000***	0.061
		(0.044)		(0.044)		

Notes: This table represents the matching exercise across the months of May and November for PNAD COVID in order to compare the matched sample used in the estimations of section 4.2 with the overall sample. Balance is tested across multiple dimensions pertaining to worker characteristics. Each dimension is regressed on a dummy that indicated whether the worker is matched or not. The coefficient and it's standard error are reported in this table. Normalised differences are also reported in this analysis. Robust standard errors in parentheses , *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

I.C.3 A Regression Discontinuity Design on the Short-term Impacts of EA

The Auxĺio Emergencial program was over a period of seven months. This rather unexpected and large rainfall of additional income can potentially have important labour market consequences on members of families receiving this aid. To test this result, it is ideal if I am able to observe the household over time. The PNAD Covid dataset is a rotational panel and so I was able to match a significant number of households in the May survey with those in the November survey. I then attempt to exploit the eligibility criteria in order to design a regression discontinuity analysis empirical strategy. This is an attempt to arrive to a causal impact of Auxĺio Emergencial, received in May, on labour market outcomes in the short term, in November. This analysis, however, remains insufficient to provide evidence on the causal effects of Auxĺio Emergencial because the time span studied is very short and the country was still struggling with the implications of Covid-19 and there are limitations for the regression discontinuity design attempted herein. Therefore, this remains additional evidence and robustness checks.

There were clear eligibility criteria for receiving Auxĺio Emergencial. I summarise those eligibility criteria across the different phases of the program in table I.18. Mainly, a maximum of two people in the household are allowed to receive the aid. People receiving the aid must be either unemployed or informally unemployed, and must not be receiving any other government support except for Bolsa Família. The household income may not exceed R$ 3,135.

If the criteria is followed accurately, then one can argue that household members just above or just below the family income cut-off of R$ 3,135 are similar on observable and, hence, unobservable characteristics. I begin by showing, in figure I.33, that there is discontinuous jump at the aforementioned cut-off in terms of an individual's probability of receiving Auxĺio Emergencial.

Since treatment was not perfect, this empirical strategy can best be described by a fuzzy regression discontinuity design. I construct the instrumental variable by assigning each individual within a family a binary variable if they satisfy all the eligibility criteria to be treated. I estimate the following equation:

I report the first stage results in table I.23. I show, across different bandwidth around the cutoff, that the first stage is strong with a high Kleiberg and Paap F statistic. However, there are certain threats to this empirical strategy. First, I need to

establish that the individuals on either side of the cut-off are similar on their observable characteristics before extending an assumption that they are also similar on their unobservable characteristics. I report the results on this balance test in table I.24. Based on the normalised differences, I find that those individuals above and below the threshold based on the bandwidth of plus and minus 140 $R, are not significantly different from one another. However, based on the t-test, I find that there are significant differences on gender, some education levels, and age. Because the latter can be subject to biases based on the sample size, I prefer the usage of normalised differences instead.

Second, for the empirical strategy to be valid, individuals should not be able to manipulate their position relative to the cut-off. This is important to test because income, my running variable, is self reported. I test this in two ways. Visually, I present a histogram of the distribution of the running variable in figure I.34. I do observe visually bunching below the threshold. Then I test if there is manipulation formally using an inference method based on based on robust bias-correction T-statistic under the null of no manipulation. q represents the power of the polynomial used in the local density estimations. I refer readers to Calonico, Cattaneo, and Titiunik (2014), Calonico, Cattaneo, and Farrell (2018) and Cattaneo, Jansson, and Ma (2017) for a detailed description of the test. I report the results in table I.25 and find that for polynomial fitting of degree 1, 2, or 3 in the unrestricted sample, and degree 2 and 3 in the restricted sample, that there is evidence of manipulation.

With this limitation in mind, causality is threatened. I implement a fuzzy regression discontinuity design (RDD) that uses a standard 2SLS procedure to estimate the impacts on the short-term November labour market of having received aid in May. I estimate the following first stage equation:

$$EA_{i,May} = \alpha + \beta\chi_i + T_{criteria} + f(income) + \epsilon_i \tag{I.C.18}$$

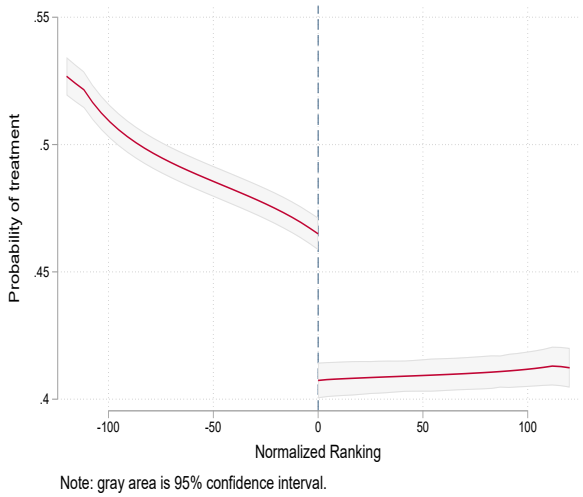
where $f(income)$ is a parametric function of the running variable, labour income. The second stage equation is estimated as follows:

$$Y_{i,Nov} = \alpha + \beta\chi_i + \lambda\chi_{i,May} + EA_{i,May} + EA_{i,Nov} + f(income) + \epsilon_i \tag{I.C.19}$$

I report the results in table I.26 across five labour market outcomes: income, hours worked, having done any form of work, being an informal employee and being in the same job location. I control for the May values of these variables as well as for

individual level characteristics in the regression. I find that having received the aid in May is positively correlated with income and the probability of being informally employed in November, but negatively correlated with the probability of having done any type of work. I note that there is no significant impact on hours worked, formality status, and being in the same job location. Further, I find that there is no statistically significant impact on health related outcomes such as displaying COVID-19 symptoms or being hospitalised if infected with COVID-19.

Figure I.33: Visual Discontinuity in the Running Variable



Notes: This figure shows the discontinuity in the probability of receiving the Auxílio Emergencial cash transfer in May against the running variable.

Table I.23: First Stage

	(1) Received EA in May	(2) Received EA in May	(3) Received EA in May	(4) Received EA in May	(5) Received EA in May	(6) Received EA in May
IV	0.0609***	0.0530***	0.0528***	0.0592***	0.0577***	0.0658***
	(0.0144)	(0.0139)	(0.0129)	(0.0123)	(0.0117)	(0.0106)
Running Variable [Income per Capita]	-0.000480	-0.000524**	-0.0000145	-0.0000635	-0.0000638	-0.0000198
	(0.000370)	(0.000242)	(0.000138)	(0.000111)	(0.0000857)	(0.0000615)
Income Range	[-40, 40]	[-60, 60]	[-80, 80]	[-100, 100]	[-120, 120]	[-140, 140]
Individual Controls	Yes	Yes	Yes	Yes	Yes	Yes
UF F.E	Yes	Yes	Yes	Yes	Yes	Yes
Kleibergen-Paap F-stat	17.9064	14.6117	16.6743	23.1953	24.2439	38.4090
Anderson-Rubin Test	0.6184	0.7531	0.8542	0.7018	0.7317	0.2206
Cragg-Donald F Stat	29.4900	26.0504	29.2727	41.3837	43.2371	70.2927

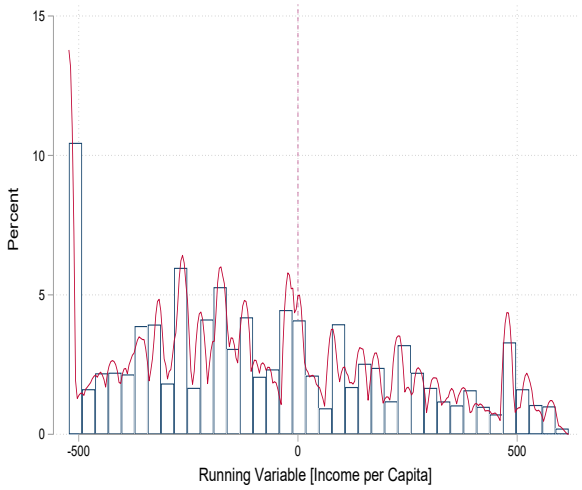
Notes: This table represents the matching exercise across the months of May and November for PNAD COVID in order to compare the matched sample used in the estimations of section 4.2 with the overall sample. Balance is tested across multiple dimensions pertaining to worker characteristics. Each dimension is regressed on a dummy that indicated whether the worker is matched or not. The coefficient and it's standard error are reported in this table. Normalised differences are also reported in this analysis. Robust standard errors in parentheses , *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table I.24: Balance Test on Observable Characteristics

Variable	(1) Above Cutoff Mean/SE	(2) Below Cutoff Mean/SE	T-test P-value (1)-(2)	Normalized difference (1)-(2)
Male	0.498	0.482	0.003**	0.034
	(0.004)	(0.004)		
Female	0.502	0.518	0.003**	-0.034
	(0.004)	(0.004)		
White	0.425	0.419	0.328	0.011
	(0.004)	(0.004)		
Black	0.096	0.095	0.637	0.005
	(0.002)	(0.002)		
Asian	0.006	0.007	0.236	-0.014
	(0.001)	(0.001)		
Brown	0.472	0.477	0.346	-0.011
	(0.004)	(0.004)		
Indigenous	0.002	0.002	0.191	-0.012
	(0.000)	(0.000)		
Without instruction	0.096	0.085	0.001***	0.039
	(0.002)	(0.002)		
Incomplete elementary school	0.305	0.298	0.152	0.016
	(0.004)	(0.004)		
Complete Elementary	0.069	0.068	0.726	0.004
	(0.002)	(0.002)		
Incomplete high school	0.108	0.098	0.005**	0.032
	(0.002)	(0.002)		
Complete high school	0.276	0.295	0.000***	-0.044
	(0.004)	(0.004)		
Incomplete higher	0.062	0.064	0.640	-0.005
	(0.002)	(0.002)		
Graduated	0.074	0.080	0.030**	-0.025
	(0.002)	(0.002)		
Postgraduate, Master's or Doctorate	0.010	0.012	0.145	-0.016
	(0.001)	(0.001)		
Age	31.797	33.718	0.000***	-0.096
	(0.158)	(0.154)		
N	15474	16190		

Notes: This table represents the matching exercise across the months of May and November for PNAD COVID in order to compare the matched sample used in the estimations of section 4.2 with the overall sample. Balance is tested across multiple dimensions pertaining to worker characteristics. Each dimension is regressed on a dummy that indicated whether the worker is matched or not. The coefficient and it's standard error are reported in this table. Normalised differences are also reported in this analysis. Robust standard errors in parentheses , *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Figure I.34: Bunching Around Threshold



Note: The figure shows a histogram of the percent of individuals within a certain window of the normalised running variable.

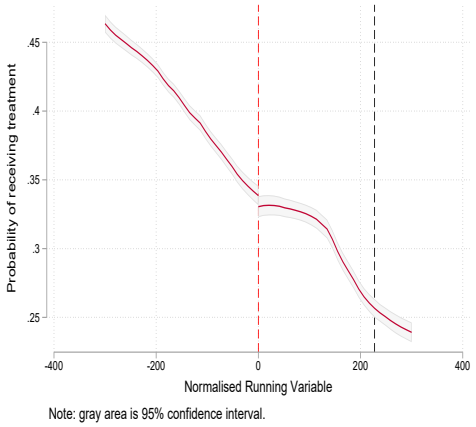
Table I.25: Manipulation Test

	Unrestricted			Restricted		
q	1	2	3	1	2	3
T statistic	7.6387	51.6821	41.2707	0.9103	-33.5719	-10.6381
p-value	0.0000	0.0000	0.0000	0.3626	0.0000	0.0000

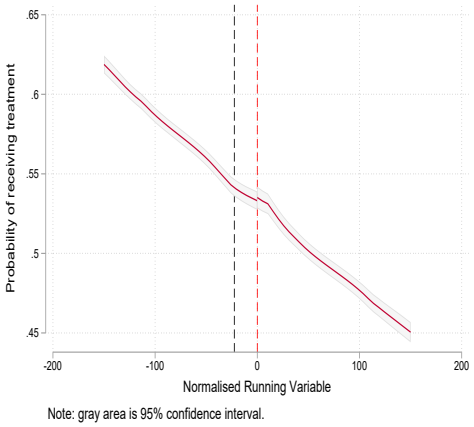
Notes: This table tests for potential manipulation at the cut-off. The inference is based on robust bias-correction T statistic under the null of no manipulation. q represents the power of the polynomial used in the local density estimations. I refer readers to Calonico, Cattaneo, and Titiunik (2014), Calonico, Cattaneo, and Farrell (2018) and Cattaneo, Jansson, and Ma (2017) for a detailed description of the test.

Figure I.35: Placebo Thresholds

(a) Threshold at 750$R



(b) Threshold at 400$R



Notes: The graphs plot the probability of having received Auxilío Emergencial in May against the running variable but at different placebo thresholds to test if there is a discontinuity at different thresholds but the one that is the true determinant of the treatment status. The grey vertical line represents the true cut-off and the red vertical line represents a placebo cut-off.

Table I.26: Results on Labour Market Outcomes in November

	(1) Log(Income)	(2) Hours Worked	(3) Done Work	(4) Informal Employee	(5) Same Job location	(6) COVID-19 Symptoms	(7) Hospital if COVID-19
Received EA in May	0.981**	3.240	-0.813***	0.199*	0.143	-0.192	-0.0313
	(0.411)	(4.837)	(0.177)	(0.115)	(0.153)	(0.147)	(0.0616)
Received EA in November	-0.652***	-2.179	0.429***	-0.0994*	-0.0723	0.0999	0.0134
	(0.216)	(2.538)	(0.0898)	(0.0580)	(0.0772)	(0.0741)	(0.0311)
Informality May	-0.216***	-1.857***	0.0339**	0.554***	-0.0104	0.0111	0.00427
	(0.0428)	(0.460)	(0.0145)	(0.0130)	(0.0128)	(0.0115)	(0.00536)
log(Income in May)	0.100***	0.00837	0.0154***	0.00801***	-0.00364	-0.000141	0.00132**
	(0.0114)	(0.0946)	(0.00231)	(0.00146)	(0.00250)	(0.00173)	(0.000658)
Hours Worked Pre-COVID	0.00650***	0.400***	-0.0120***	-0.000368	-0.00234***	0.00175***	0.000485***
	(0.00144)	(0.0178)	(0.000524)	(0.000255)	(0.000485)	(0.000362)	(0.000134)
Hours Worked in May	0.00114	0.215***	0.0186***	0.000568***	0.00457***	-0.00105***	-0.000334***
	(0.000754)	(0.00960)	(0.000290)	(0.000145)	(0.000437)	(0.000224)	(0.000106)
Individual Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
UF F.E	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Kleibergen-Paap F-stat	34.9093	35.6858	30.8875	30.8875	20.9046	30.8875	30.8875
Cragg-Donald F Stat	63.1805	64.7622	55.6667	55.6667	38.7416	55.6667	55.6667

Notes: This table represents the matching exercise across the months of May and November for PNAD COVID in order to compare the matched sample used in the estimations of section 4.2 with the overall sample. Balance is tested across multiple dimensions pertaining to worker characteristics. Each dimension is regressed on a dummy that indicated whether the worker is matched or not. The coefficient and it's standard error are reported in this table. Normalised differences are also reported in this analysis. Robust standard errors in parentheses , *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

I.D Supplementary Material: Containment Measures

In section 4.1, I merge individuals across two quarters in PNAD Continua. I merge along the following variables: day, month and year of birth, gender, race, the primary sampling unit, role in the household, household survey selection number, Federation Unit and urban or rural regions. However, due to the rotational nature of the survey, not all the individuals across the two quarters. Since my analysis is done on the merged sample only, it is important to understand how this matched sample differs from the unmatched sample on observable characteristics to be able to make a statement on whether the analysis is representative of all individuals. I consider the following observable characteristics: age, gender, race, education, firm size, family structure, status of informality and sector of occupation. I report the t-test p-value as well as the normalised difference. I find that, based on the normalised differences, there is no statistically significant difference across the matched and unmatched sample on observable characteristics.[47]

Table I.27 – Balance Tests for PNAD Continua

	(1)		(2)		T-test	Normalized
	Unmatched Sample		Matched Sample		P-value	Difference
Agri	103056	0.096	75947	0.099	0.043**	-0.008
		(0.001)		(0.001)		
Natural Resources	103056	0.012	75947	0.012	0.292	-0.005
		(0.000)		(0.000)		
Manuf.	103056	0.119	75947	0.121	0.209	-0.006
		(0.001)		(0.001)		
Construct.	103056	0.079	75947	0.076	0.011**	0.013
		(0.001)		(0.001)		
Trade	103056	0.196	75947	0.187	0.000***	0.024
		(0.001)		(0.001)		

Continued on next page

[47] The normalised differences measure is preferred to the T-test P-value because the latter can be highly sensitive to the sample size. I follow Imbens and Wooldridge (2009) and consider significant differences as those with a normalised difference exceeding 0.25.

Table I.27 – *Continued from previous page*

	(1)		(2)		T-test	Normalized
	Unmatched Sample		Matched Sample		P-value	Difference
Transport	103056	0.052	75947	0.050	0.052*	0.010
		(0.001)		(0.001)		
Hospitality	103056	0.056	75947	0.049	0.000***	0.033
		(0.001)		(0.001)		
Info & Com.	103056	0.016	75947	0.017	0.085*	-0.010
		(0.000)		(0.000)		
Finance	103056	0.014	75947	0.016	0.001***	-0.019
		(0.000)		(0.000)		
Real Estate	103056	0.007	75947	0.006	0.469	0.004
		(0.000)		(0.000)		
Admin., Prof. & Serv.	103056	0.129	75947	0.129	0.886	-0.001
		(0.001)		(0.001)		
Public Sector	103056	0.049	75947	0.057	0.000***	-0.036
		(0.001)		(0.001)		
Education	103056	0.057	75947	0.064	0.000***	-0.028
		(0.001)		(0.001)		
Healthcare	103056	0.059	75947	0.062	0.058*	-0.010
		(0.001)		(0.001)		
HH workers	103056	0.060	75947	0.056	0.001***	0.016
		(0.001)		(0.001)		
Informal	249623	0.215	185199	0.212	0.049**	0.006
		(0.001)		(0.001)		
Male	249623	0.491	185199	0.486	0.006**	0.008
		(0.001)		(0.001)		

Continued on next page

Table I.27 – *Continued from previous page*

	(1)		(2)		T-test	Normalized
	Unmatched Sample		Matched Sample		P-value	Difference
Firm size ≤	83676	0.614	60728	0.634	0.000***	-0.042
10 employees		(0.002)		(0.002)		
Firm size btw 10	83676	0.118	60728	0.111	0.000***	0.021
and 50 employees		(0.001)		(0.001)		
Firm size > 50 employees	83676	0.268	60728	0.255	0.000***	0.032
		(0.002)		(0.002)		
No education or<	235017	0.082	175504	0.072	0.000***	0.032
1 year of study		(0.001)		(0.001)		
Incomplete elementary	235017	0.330	175504	0.329	0.362	0.003
		(0.001)		(0.001)		
Complete elementary	235017	0.078	175504	0.077	0.492	0.002
		(0.001)		(0.001)		
Incomplete high school	235017	0.073	175504	0.070	0.000***	0.012
		(0.001)		(0.001)		
Complete high school	235017	0.261	175504	0.258	0.008**	0.009
		(0.001)		(0.001)		
Incomplete	235017	0.046	175504	0.048	0.010**	-0.009
higher education		(0.000)		(0.001)		
Graduated	235017	0.129	175504	0.146	0.000***	-0.052
		(0.001)		(0.001)		
14-34	204584	0.416	153520	0.388	0.000***	0.059
		(0.001)		(0.001)		
35-53	204584	0.328	153520	0.339	0.000***	-0.023
		(0.001)		(0.001)		

Continued on next page

Table I.27 – *Continued from previous page*

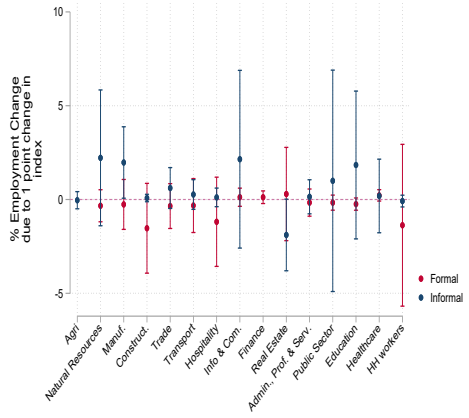
	(1)		(2)		T-test	Normalized
	Unmatched Sample		Matched Sample		P-value	Difference
54+	204584	0.256	153520	0.274	0.000***	-0.037
		(0.001)		(0.001)		
White	249623	0.423	185199	0.446	0.000***	-0.047
		(0.001)		(0.001)		
Black	249623	0.099	185199	0.087	0.000***	0.045
		(0.001)		(0.001)		
Asian	249623	0.007	185199	0.006	0.000***	0.013
		(0.000)		(0.000)		
Brown	249623	0.466	185199	0.458	0.000***	0.016
		(0.001)		(0.001)		
Indigenous	249623	0.004	185199	0.003	0.000***	0.016
		(0.000)		(0.000)		

Table I.27: Balance Tests for PNAD Continua Matched Sample

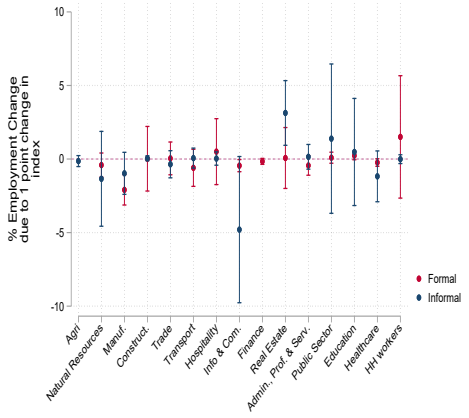
Table III.32represents the matching exercise across two survey quarters in 2021 in order to compare the matched sample used in the estimations of section 4.1 with the overall sample. Balance is tested across multiple dimensions pertaining to worker characteristics. Each dimension is regressed on a dummy that indicated whether the worker is matched or not. The coefficient and it's standard error are reported in this table. Normalised differences are also reported in this analysis.

Figure I.36: Percent Change in Formal and Informal Employment by Sector in Response to Government Measures

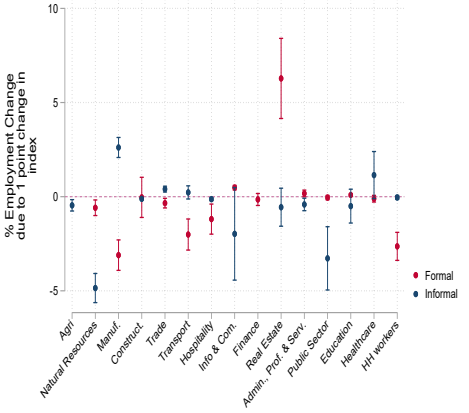
(a) Economic Support Index



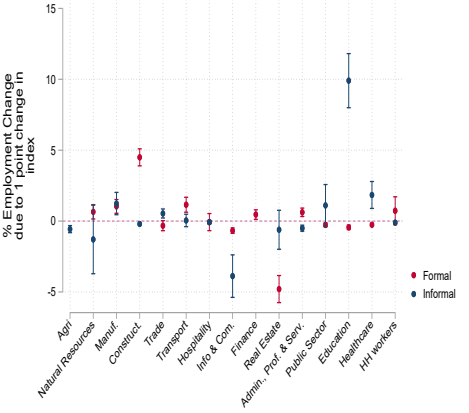
(b) L1. Economic Support Index



(c) Stringency Index



(d) L1. Stringency Index



Notes: The graphs show the point estimates and confidence intervals of $\theta_{1,s}$ and $\theta_{2,s}$ in equation (4.15) for each government response measure (Stringency Index, and Economic Support Index), by formality status. Formal employment is omitted in the Agriculture sector and informal employment is omitted in the Finance sector. The reason is because each of these employment is small and hence the confidence intervals are very large indicating that they are not being estimated with precision. Those estimates, which are insignificant from zero, distort the scale when included. The equation is estimated for each one of the sectors on the x-axis. Source: PNAD-Continua and author's estimations. 90 percent confidence intervals are depicted around the point estimates.

In what follows, I present summary statistics on the governmental containment measures, averaged across all years for each Federation Unit. The data presented is based on the Oxford Covid-19 Government Response Tracker (OxCGRT). Table I.28 shows the average of the aggregate of the main aggregate containment measures: Stringency Index, Government response index, Health Containment measures, Economic Support Index.

The containment measures recorded information on containment policies adopted

by a certain country pertaining to the closure of schools or various restrictions on mobility. it also includes health related measures such as contact testing and tracing as well as short-term investments in health care and vaccines. I show the average values across Federation Units of the subcategories of this measure in table I.29. The Economic Support Index shows the economic policies implemented by a certain country, taking into consideration income support or debt relief. I show the average values of the subcomponents of this index in table I.30. The Stringency Index shows the extent of the lockdown measures taken by a certain country. The subcomponents of this measure as well as their average value in table I.31. For a more detailed description on each of the measures, I refer readers to the methodology on these policy indices by Oxford COVID-19 Government Response Tracker..

Table I.28: Average of Main Aggregate Indices across Time by Federation Units

	Mean Across Years			
	Stringency Index	Government Response Index	Health Containment Measures	Economic Support Index
Federation Unit				
Acre	48.3	38.9	43.1	11.8
Alagoas	48.5	41.8	45.9	15.2
Amapa	50.4	40.1	44.3	12.5
Amazonas	53.2	43.5	46.8	21.9
Bahia	47.6	41.3	45.3	15.3
Ceara	50.3	45.6	49.3	20.7
Distrito Federal	36.7	38.7	41.7	17.8
Espirito Santo	46.3	41.5	44.0	25.8
Goias	44.7	39.4	43.3	12.8
Maranhao	39.0	38.0	40.5	20.8
Mato Grosso	47.4	39.8	43.8	13.5
Mato Grosso do Sul	45.5	39.6	42.6	20.3
Minas Gerais	50.2	42.7	46.0	19.7
Para	45.7	37.1	40.8	15.4
Paraiba	44.8	39.8	43.5	15.5
Parana	52.2	51.1	53.1	38.3
Pernambuco	46.1	41.0	45.1	13.9
Piaui	42.9	40.9	44.8	15.4
Rio Grande do Norte	46.5	39.6	42.8	17.6
Rio Grande do Sul	52.4	46.4	51.3	13.1
Rio de Janeiro	44.0	38.9	40.3	31.7
Rondonia	57.7	44.6	48.1	21.2
Roraima	51.1	42.8	45.5	25.7
Santa Catarina	47.3	39.8	43.9	12.3
Sao Paulo	42.4	40.1	43.3	18.3
Sergipe	47.5	41.1	43.4	26.3
Tocantins	49.9	41.5	45.3	16.6
Brazil	45.3	43.8	45.8	31.0

Notes: The table shows the average values on each indicator across time since first implemented. The table shows the policy values on the Stringency Index, which measures the strictness of lockdown, on the health containment index which measures testing policies, closures of venues, and investments in health care. The Economic Support Index measures income support and debt relief, and the government response index, which is an overall measures of the response of the government across all indicators. Higher values on each index indicates greater strictness. Source: Oxford COVID-19 Government Response Tracker.

Table I.29: Average of the Subcomponents of the Containment Measure across Time by Federation Units

	Mean Across Years							
	Schools	Workplace	Public Events	Gatherings	Public Transport	Stay Home	Internal Movement	International Travel
Federation Unit								
Acre	2.3	1.4	1.3	2.1	0.8	1.0	0.9	0.6
Alagoas	1.9	1.6	1.4	2.5	0.6	0.9	0.7	0.8
Amapa	2.4	1.7	1.6	1.6	0.6	1.3	1.1	0.7
Amazonas	2.0	1.6	1.6	2.3	0.5	1.3	1.3	0.8
Bahia	2.1	1.6	1.6	1.6	0.7	0.8	0.8	0.5
Ceara	1.9	1.2	1.6	2.5	0.7	1.2	0.7	0.7
Distrito Federal	1.9	1.2	1.7	0.0	0.1	1.1	0.0	0.2
Espirito Santo	1.9	1.8	1.5	2.4	0.5	1.0	0.6	0.6
Goias	1.9	1.7	1.5	1.7	0.7	0.9	0.1	0.5
Maranhao	1.8	1.3	1.2	1.4	0.5	0.8	0.3	0.6
Mato Grosso	2.2	1.2	1.5	1.6	0.8	1.3	0.9	0.8
Mato Grosso do Sul	1.9	1.3	1.3	2.0	0.6	1.5	0.5	0.6
Minas Gerais	2.3	1.7	1.3	2.3	0.7	0.8	0.9	0.5
Para	1.7	1.6	1.6	2.5	0.3	1.0	0.7	0.7
Paraiba	2.1	1.4	1.2	1.6	0.8	0.9	0.8	0.6
Parana	2.0	1.8	1.6	2.7	0.5	1.2	0.4	0.8
Pernambuco	2.1	1.2	1.3	2.4	0.3	0.5	1.1	0.7
Piaui	1.9	1.3	1.4	1.5	0.6	0.9	0.6	0.7
Rio Grande do Norte	2.2	1.6	1.5	1.1	0.6	1.0	0.8	0.8
Rio Grande do Sul	1.8	1.6	1.6	3.0	0.8	0.9	0.6	0.5
Rio de Janeiro	2.1	1.4	1.4	2.0	0.8	0.3	0.5	0.6
Rondonia	2.2	1.8	1.4	2.8	0.6	1.1	1.2	1.8
Roraima	2.0	1.6	1.5	2.5	0.7	1.1	1.2	0.7
Santa Catarina	1.5	1.3	1.4	2.5	0.8	0.6	1.1	0.6
Sao Paulo	1.7	1.7	1.6	1.6	0.3	0.9	0.2	0.8
Sergipe	1.9	1.5	1.6	2.6	0.7	0.8	0.2	1.1
Tocantins	1.7	1.6	1.6	2.8	0.7	0.9	0.9	0.5
Brazil	1.3	1.6	1.0	1.8	0.7	0.9	1.7	1.5

Notes: The table shows the average values on each indicator across time since first implemented. The indicators represent the subcomponents on the Containment measure aggregate policy index and reflect the extent of closure of schools, workplace, public events, public transport, or stay at home measures and restrictions on internal and international movement. Higher values on each index indicates greater strictness. Source: Oxford COVID-19 Government Response Tracker.

Table I.30: Average of the Subcomponents of the Economic Support Measure across Time by Federation Units

	Mean Across Years	
	Income Support	Debt Relief
Federation Unit		
Acre	0.2	0.3
Alagoas	0.3	0.4
Amapa	0.2	0.3
Amazonas	0.4	0.6
Bahia	0.3	0.4
Ceara	0.4	0.5
Distrito Federal	0.0	0.7
Espirito Santo	0.9	0.4
Goias	0.3	0.3
Maranhao	0.4	0.5
Mato Grosso	0.3	0.3
Mato Grosso do Sul	0.2	0.6
Minas Gerais	0.4	0.5
Para	0.3	0.4
Paraiba	0.2	0.4
Parana	0.4	1.3
Pernambuco	0.3	0.4
Piaui	0.2	0.4
Rio Grande do Norte	0.3	0.5
Rio Grande do Sul	0.4	0.3
Rio de Janeiro	0.4	0.9
Rondonia	0.4	0.5
Roraima	0.3	0.8
Santa Catarina	0.3	0.2
Sao Paulo	0.4	0.3
Sergipe	0.7	0.4
Tocantins	0.3	0.4
Brazil	0.7	0.7

Notes: The table shows the average values on each indicator across time since first implemented. The indicators represent the subcomponents on the Economic support measure aggregate policy index and reflect the extent of income support and debt relief. Higher values on each index indicates greater strictness. Source: Oxford COVID-19 Government Response Tracker.

Table I.31: Average of the Subcomponents of the Stringency Measure across Time by Federation Units

	Mean Across Years					
	Policy	Info Campaigns	Testing Policy	Contract Tracing	Facial Covering	Protect Elderly
Federation Unit						
Acre	1.5	1.4	0.5	0.7	3.1	0.2
Alagoas	1.5	1.6	1.1	0.8	3.0	1.0
Amapa	1.3	1.5	0.7	0.7	2.8	0.5
Amazonas	1.5	1.8	1.2	0.8	2.4	0.2
Bahia	1.4	1.8	1.1	0.7	3.1	0.9
Ceara	1.5	1.7	1.4	1.0	3.1	1.1
Distrito Federal	1.5	1.9	1.6	0.2	3.1	2.5
Espirito Santo	1.4	1.6	1.2	0.9	2.8	0.4
Goias	1.4	1.7	1.6	0.7	3.1	0.4
Maranhao	1.3	1.7	0.8	0.8	3.0	1.4
Mato Grosso	1.2	1.7	1.2	0.7	3.2	0.3
Mato Grosso do Sul	1.4	1.7	1.2	0.7	2.2	1.0
Minas Gerais	1.2	1.7	1.2	0.7	2.6	1.0
Para	1.3	1.4	0.5	0.6	2.9	0.3
Paraiba	1.3	1.7	1.0	0.7	2.7	1.3
Parana	1.4	1.9	1.8	1.0	3.2	1.7
Pernambuco	1.5	1.7	1.4	0.7	2.5	1.2
Piaui	1.2	1.7	0.9	1.3	2.8	1.6
Rio Grande do Norte	1.4	1.7	1.3	0.6	2.7	0.5
Rio Grande do Sul	1.3	1.8	1.5	1.2	2.7	1.4
Rio de Janeiro	1.4	1.7	0.9	0.5	3.1	0.4
Rondonia	1.4	1.8	0.5	0.7	2.7	0.3
Roraima	1.4	1.3	0.8	0.8	2.6	0.5
Santa Catarina	1.2	1.7	1.0	0.8	2.8	0.8
Sao Paulo	1.4	1.8	2.0	0.7	3.0	0.7
Sergipe	1.5	1.6	1.2	0.4	3.0	0.4
Tocantins	1.3	1.7	0.8	0.8	2.9	0.6
Brazil	1.8	1.9	1.8	1.1	3.1	0.7

Notes: The table shows the average values on each indicator across time since first implemented. The indicators represent the subcomponents on the Economic support measure aggregate policy index and reflect the extent of the lockdown style. It includes information campaigns, policies pertaining to testing and contact tracing and protecting the elderly, as well as policies on facial coverings. Higher values on each index indicates greater strictness. Source: Oxford COVID-19 Government Response Tracker.

CHAPTER II

Immigration and Trade in a Middle-Income Country

Abstract Using new data on migration flows, I explore the impact of international immigration on trade, exploiting the spatial variation across Brazilian municipalities from 2010 to 2017. First, I discuss the preference and fixed costs underlying channels through which immigration affects trade based on the Tai (2009). Second, I estimate the correlation between immigration and trade using a rich set of fixed effects and motivated by an augmented gravity model. And third, I provide a causal estimate of the impact at the municipality as well as the country-municipality levels. In order to account for the endogeneity of migrant locations within the country and the timing of immigration, I use the commonality of language between early settlements of migrant groups and new migration inflows in recent years. Based on the causal estimates at the municipality level, I find that doubling the average of overall migrant share of the population across municipalities increases differentiated imports by 0.022%. Further, I present evidence that the impact of immigration differs depending on characteristics of the immigrant's country of origin. I find that immigrants arriving from countries with higher levels of interpersonal trust and lower level of ethnic fractionalisation cause an increase in differentiated as well as homogeneous imports. Whereas, immigrants from less developed economies have a positive impact on differentiated imports, indicating that immigrants' specialised knowledge might be able to substitute weak institutions in their home country. The paper highlights the pro-trade effects of immigrants, in a middle-income country, and presents evidence that this impact is not only

heterogeneous depending on the goods' level of differentiation, but also depending on characteristics of the migrants' country of origin.

1 Introduction

There is a substantial body of evidence on the impact of immigration on trade. Most studies find that migration from one country to another tends to lead to an increase in trade between those two countries (Parsons and Winters, 2014). There are multiple explanations for this finding. One reason might be that migrants can increase the level of information about the country of origin (fixed trade cost channel), hence reducing the costs involved with trade. Another reason might be that migrants can demand products from the home country for their own consumption, perhaps in the process shaping the preferences of other residents in the destination country (preferences channel). A common feature of the literature on this topic shares a focus on South-North or North-North migration; that is, between low or middle-income and high-income countries or between high-income countries. However, South-South migration flows form an increasingly significant portion of global flows; in 2015, South-South migration flows was estimated at 90.2 million in comparison with 85.3 million for South-North migration (International Organization for Migration, 2017).

In this study, I depart from that focus and explore the impacts of immigration on trade in the context of a middle-income country. In particular, I ask: How does international immigration impact imports and exports in a middle-income economy? This is important to study because the sectoral composition, informality levels, policies of integration, and immigrant education profile would differ between middle-income and high-income economies, and hence the impacts of immigration on trade may not be the same in those contexts. In order to estimate the impact of international immigration on trade, I focus this study on Brazil, and I base my analysis on the spatial variation in the location of immigrants and trading firms between 2010 and 2017. By studying the impact at different degrees of product differentiation, I present suggestive evidence that points to potential channels of impact, whether through changes in preferences or fixed costs, or both. I also study the impacts at the product extensive margin, and present evidence of heterogeneous effects depending on certain characteristics of the country of origin of the immigrant.[1]

Brazil, the wealthiest economy in Latin America, offers a good context in which to explore questions related to trade and immigration for the following reasons.[2] First, with the strategic global positioning of Brazil on the international political sphere and

[1]Products at the extensive margin are defined as all products that are traded for the first time between a certain municipality and country pair between 2010 and 2017.

[2]Brazil is the wealthiest economy in Latin America in terms of total annual GDP.

its hosting of major sporting events such a the 2014 FIFA world cup and 2016 Olympics and Paralympics (Uebel, 2020), immigration rose significantly in the past decade.[3] The number of migrants that arrived into Brazil between 2010 and 2017, which is the period I study in this paper, represented an eight-fold increase from the pre-2000 levels. Second, Brazilian municipalities had different exposure to migrants. Some municipalities did not experience much growth in the migrant population, while the share of the migrant population increased substantially in others, in the sample period. This variation in exposure to migration shocks across municipalities will have an impact on the choices that firms undertake including import and export decisions, and the presence of immigrants might affect trade differently depending on the elasticity of substitution of the traded product. Third, Brazil receives a diverse set of migrants that differ in terms of their cultural proximity to Brazilians. Immigration to Brazil still accounts for a small, but growing, share of the country's population and my analysis can inform future policymaking in the country as the migrant population continues to increase. Fourth, Brazil has excellent data on trade that varies at the municipality, product and trading country levels. It is also further disaggregated at the Harmonised Coding and Description System (HS-4) four digit product code level which allows me, through a series of cross-walks, to combine it with data on product elasticities of substitution.

I first approach this research question by briefly discussing the potential channels through which immigration impacts trade based on Tai (2009). I provide an empirical link of the suggested channels to the data, and I present a hypothesis on the underlying mechanisms: immigrants can potentially have the highest impact on preferences for the highly differentiated goods, and the highest reduction in fixed costs for the less differentiated ones, due to the more straightforward transaction costs for the latter.[4] Building on the model, I then proceed to provide empirical evidence on the relationship between immigration and trade. I begin by estimating the relationship, at the country, year, and municipality levels, motivated by the augmented gravity equation based on

[3]Brazil's foreign policy in the past decade contributed to its international presence. Part of this foreign policy is the *Política Externa Migratória Brasileira (PEXM)* that influenced immigration into Brazil from Latin America and West Africa. As part of PEXM, there has been certain strategic policies that aimed to promote the country abroad such as 1- participation in humanitarian missions in Haiti and some African countries, scholarships for individuals from Latin America and Africa, The Mais Médicos (More Doctors) Program, and its presence in the major sporting events such a the 2014 FIFA world cup and 2016 Olympics and Paralympics (Uebel, 2020; Moreira, 2015).

[4]However, the impact on these channels will translate to different final effects on trade. The final effect will be mediated by, among others, the extent of which new firms will start to trade (firm extensive margin), and the respective increase in market share of differentiated versus homogeneous goods.

Tai (2009). I estimate this empirical analysis using Ordinary Least Squares (OLS) and Psuedo Poisson Maximum Likelihood (PPML) with a rich and demanding set of fixed effects that help control for important variables that could affect the relationship such as common language or borders, distance, trade agreements, colonial relationship, among others (Anderson and van Wincoop, 2003; Feenstra, 2003; Helpman et al., 2008; Head and Mayer, 2014). However, despite the rich set of fixed effects employed in the estimation, it remains subject to endogeneity concerns because immigrants do not locate randomly within a host country. They are likely to choose the locations that are favourable to them, economically, socially, politically, and otherwise. Also, a concern of reverse causality persists, as trade itself could be a catalyst for more immigration. Therefore, other studies on the impacts of immigration, including those studying the impacts on trade such as Peri and Requena-Silvente (2010) and Ehrhart et al. (2014), use an instrumental variable approach to identify the causal impact of immigration. In order to account for this endogeneity, I use a shift-share instrumental variable approach.

My instrumental variable approach exploits the language commonalities between newly arriving immigrants and historic immigration settlements. For this approach to have plausible exogeneity, the necessary assumption is that earlier immigration settlements located in municipalities due to factors that do not affect current trade dynamics. It could remain the case that earlier migrant settlements established certain trade relations that persist across time and remain correlated with future migrant flows. Further, one would worry that trade itself drives more immigrants into the country, a channel that poses a clear challenge to my claims of causality. I run a series of robustness tests to ameliorate these concerns that could violate the exogeneity assumption. For example, I test for correlation between historic language shares and pre-trends in trade, and for correlation with baseline municipality characteristics that could have a long term effect on trade. I also study whether immigration had a pro-trade creation effect at the product-extensive margin, where immigrants have the potential of stimulating trade in new products. As an additional robustness check, I perform the analysis including only the countries that had no or less than 15 immigrants present in 2000, but had much larger and recent flows within the period of analysis, from 2010 to 2017. I find results, for this subset of countries, that are consistent with those observed for the full sample of countries.

The results of this paper are as follows. First, based on the estimated correlation at the municipality, country, and year levels, and using a restrictive set of fixed effects,

I find that immigration is associated with an increase in exports at low levels of differentiation, and an increase in imports at low and high levels of differentiation. Both of these effects are reported at the product extensive margin. Upon estimating a causal impact at a municipality level, I find that the main impact of immigration is through an increase in differentiated imports. In particular, an equivalent to doubling the average of overall migrant share to the population across municipalities would result in an increase in differentiated imports by 0.022%. This result is fairly consistent with the causal estimation at the country and municipality levels, where I find an increase in differentiated and homogeneous imports at the extensive margin, and for imports at low to mid levels of differentiation as classified by Broda and Weinstein (2006)'s sector elasticity of substitution. Second, the presence of an impact on more differentiated imports rather than exports, is suggestive that the preferences channel is more prominent than the fixed cost channel. Third, I show that there is significant heterogeneity in the impact of immigrants on trade, not only based on the level of product differentiation, but also depending on the country of origin characteristics. I find that immigrants arriving from countries with lower levels of ethnic fractionalisation and higher levels of interpersonal trust cause an increase in differentiated and homogeneous imports. Interestingly, immigrants from less developed economies, by GDP levels, have a positive impact on differentiated imports, indicating that, in addition to demanding familiar varieties, immigrants' specialised knowledge might be able to facilitate trade and substitute weak institutions in their home country, a result that is consistent with some of the literature on the subject (Ehrhart et al., 2014).

This paper has several contributions. First, I explore this research question in the context of South-South migration. At a global level, South-South migration accounts for larger flows of people than South-North movements (International Organization for Migration, 2017). The impact of immigration on trade can be distinct in this context due to differences in the nature of comparative advantage and endowments, and to differences in policies that the Global South adopts for the integration of migrants. These factors could affect the economic integration of newcomers and their interaction with the host communities. Second, I present a different way to think about the past settlement shift-share Bartik instrument in order to arrive at a causal estimate of the impact of migration on trade. Brazil received immigrants from many new countries that were not present historically. So in order to predict current migrant locations, I use the language spoken by the majority of the population in the country as the common factor between historical locations and new immigrants. The underlying hypothesis, is

that immigrants locate near others who are culturally similar to them. This broader definition encompasses locating near other immigrants that share a common language, not just a common country of origin. Third, I show evidence that the impact on trade varies depending on the traded goods' level of differentiation, and on the migrants' country of origin characteristics.

The rest of the paper is organised as follows. Section 2 describes the background of immigration into Brazil and a literature review on the links between immigration and trade. Section 4 describes the data and provides definitions on what is meant by the different disaggregated traded goods. Section 4 presents a discussion of the potential channels of impact and the results from the first estimation strategy. Following this, I estimate a causal impact of immigration on trade at the municipality level as well as municipality and country levels. Section 5 discusses the identification strategy and section 5.2 presents and discusses the results. Section 6 discusses robustness checks and the limitations of the identification strategy. Section 7 concludes and provides policy implications.

2 Migration and Trade

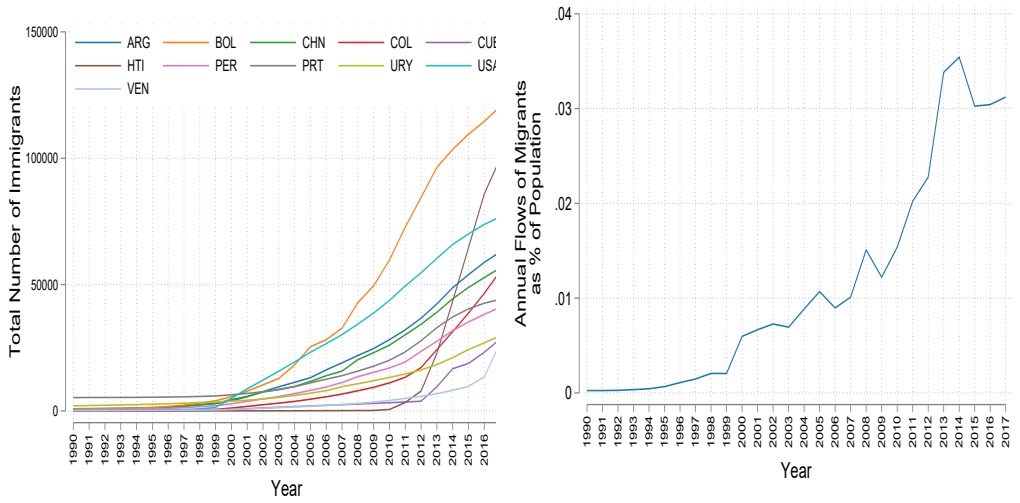
Modern Brazil has not been a traditional destination for immigrants. This has started to change in recent years, particularly because of the country's strategic positioning as a regional power in the world and because of receiving new migrant inflows from countries such as Haiti and Venezuela. In the latter cases, the inflows responded mainly to exogenous events in the countries of origin and, as shown in figure II.1(a), inflows from these countries increase from a less than a few hundred to over 10,000 per year relatively quickly. Also, there has been a steady migrant inflow from other countries such as Bolivia and the United States. My analysis studies the impact of the overall immigrant inflow during 2010-2017 on trade. Hundreds of thousands of immigrants came to Brazil during that period, with immigrants from most of the world's countries. While the number is large overall, it is fairly small relative to the overall population of the country (less than 1%). The distribution of immigrants across Brazilian municipalities is depicted in the maps in figure II.2.

Empirically, there is a large body of literature that demonstrates the relationship between migration and trade. Gould (1994) is considered a pioneering study in this area. The author uses U.S. data from 1970-1986 to study the link between migration

Figure II.1: Immigration by Year

(a) Accumulated Flows By Country of Origin

(b) Annual Flows as a % of Total Population

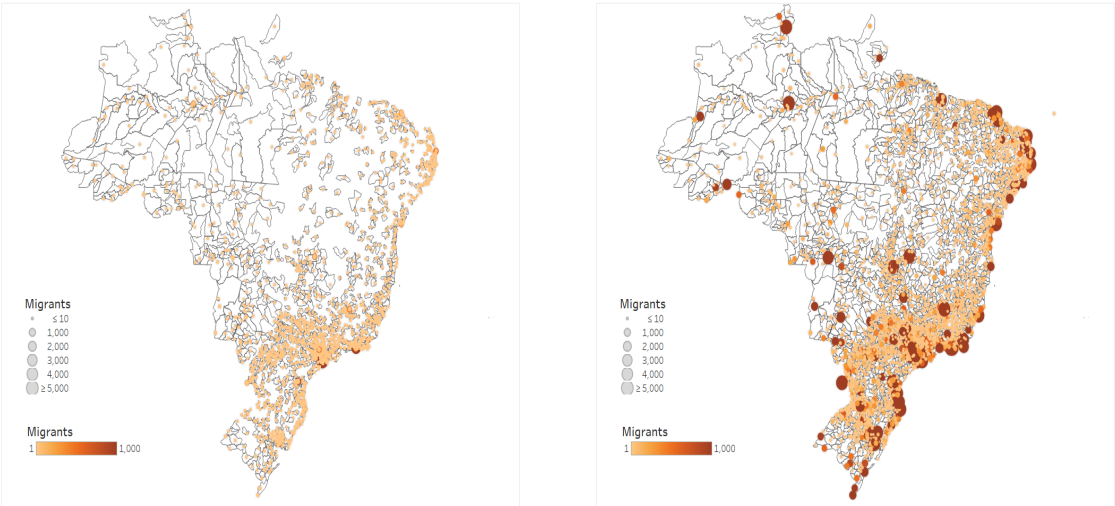


Notes: Figure II.1(a) plots the accumulated flows from a certain country. The countries that sent the most immigrants (at least 4000 immigrants across all years) into Brazil are shown in this graph. Figure II.1(b) plots the immigration annual flows into Brazil as a percent of the total population by year. The year on the x-axis represents the year of arrival of migrants into the country. Author's analysis of SINCRE dataset.

Figure II.2: Distribution of All Migrants in Brazil

(a) Before 2000

(b) Between 2010 and 2017



Notes: These maps show the spatial distribution of immigrants across municipalities in Brazil before and after 2009. Location data is from IBGE statistics. Author's analysis of SINCRE dataset.

and trade with 47 high income countries. Gould (1994) considers two basic mechanisms that could create links between trade and migration. First, immigrants bring specialised information about their home market that helps lower transaction costs, and second, immigrants bring with them preferences for certain home country products. The author finds that immigration has a stronger impact on exports than on imports, pointing perhaps to the presence of the former channel than the latter one. Similarly, Girma and Yu (2002) find that migration affects exports more than imports in the context of the United Kingdom and its trade with 48 countries from 1981 - 1993. In both studies, the sample used for the analysis consists mainly of high-income countries. In contrast, Head and Ries (1998) and Wagner et al. (2002) look at a larger sample of countries (136) in the Canadian context and find that migration impacts imports more than exports.

Hence, the literature has been pointing to two key direct channels by which migrants affect trade between their home and host countries. The first channel relates to immigrants' possession of information about markets in their country of origin. This information can overcome fixed costs associated with trade. It includes the migrants' impacts on language, culture and other barriers to foreign trade, their supply of new information, and their ability to connect businesses and suppliers across countries (Combes et al., 2005, Dunlevy, 2006, Felbermayr and Jung, 2009, Felbermayr et al., 2010, Herander and Saavedra, 2005, Rauch and Trindade, 2002). The effect of this channel is magnified when the two countries have significant differences in culture, customs, institutions and languages. Moreover, the information supplied by migrants becomes more important for trade with low- and middle-income countries, as in my case study, since experience and knowledge play a greater role in assessing the trustworthiness of the trading parties (Strozzi, 2017).

The second channel relates to preferences about different products. It embodies the enhancement of trade caused by the migrants' knowledge of different products from their home country (Girma and Yu, 2002, Gould, 1994, Head and Ries, 1998, Wagner et al., 2002). This difference could relate to factors such as quality or presence of new products. The preference channel also includes the increase of the host countries imports caused by the migrants' demand of goods produced in their home countries and a possible shift in the taste of the host country's residents. However, this increase could be dampened because as the number of migrants increases, local firms could start manufacturing these goods (Strozzi, 2017). Further, the basis of the preference mechanism is that immigrants demand more of their home country products. In the

Brazilian case, immigration is very small relative to the population and hence, relying only on the personal consumption of immigrants, the preference channel may not affect the levels of differentiated trade. What could be a stronger effect, is actually the diffusion of preferences from the immigrants to the natives, or "cultural transmission" (Tai, 2009). Such a phenomenon has the potential to affect an entire community as opposed to being limited to immigrant themselves. The model of Tai (2009) that I use to motivate and ground the findings allows for this wider definition of preferences.

It is also possible for migration to affect homogeneous and differentiated goods differently. Homogeneous goods are those for which consumers may neither know nor care about the source of the product they are purchasing (e.g., corn, oil, wheat, etc.). Differentiated goods are products with many characteristics that vary across suppliers and may even be specifically tailored to the end-user's needs. Auto-mobiles are perhaps the most often cited example of such a good; in fact, most consumer goods (e.g., toys, apparel, cookware) are classified as differentiated. For products which are homogeneous, the preferences channels is unlikely to play a major role. On the other hand, these preference effects can increase demand and imports for differentiated goods (Rauch and Trindade, 2002). Some results have shown that indeed migrants do affect differentiated products disproportionately. For instance, Peri and Requena-Silvente (2010) look at the Spanish context and find that doubling the number of immigrants increases exports by 10% and this effect is higher for differentiated goods. Interestingly, Briant et al. (2014) find that migration is more important for differentiated goods and matters for simple products when institutional quality in the country of origin is poor. Ehrhart et al. (2014) find that migrants can enhance trade with countries where low institutional quality might otherwise serve as a barrier (e.g., because of issues around contract enforcement). They also find that the link between migrants and trade is particularly important for differentiated products.

This paper studies the impact on trade due to international immigration in a middle-income country context. Assunção and Verdier (2020) present work that is closest to my analysis. Although, the authors estimate the analysis at the municipality level, they use firm level data to link workers to exporting and importing firms. However, the authors do not have data on the destination or origin of the firm's imports or exports nor do they have data on all the specific nationalities of immigrants (so they only focus on 30 countries). In contrast, I match datasets at a municipality level that allow me to study the universe of international immigrants in Brazil, and link the country of origin of the migrant with the country that a certain municipality

is trading with. Assunção and Verdier (2020) estimate an augmented gravity model, which this paper also estimates but I use a more restrictive set of fixed effects and I go beyond that to estimate a causal impact of immigration on trade using a shift-share IV. Assunção and Verdier (2020) are interesting, as they find that immigrants are able to have pro-trade effects for differentiated exports and imports at the intensive and extensive margins. They also find evidence of heterogeneous impacts based on institutional development or cultural proximity in the country of origin. Another interesting paper that links to my research is Pellegrina and Sotelo (2021)'s analysis of the internal migration of farmers to the west of Brazil between 1950 and 2010, and document how farmers, through reshaping comparative advantage, were able to contribute to Brazil's rise as a commodity exporter. This paper doesn't discuss the channel of potential alternation in the comparative advantage across regions due to immigration and the worker's original occupation, but is a promising avenue to explore further.

In figure II.3(a), I show the evolution of Brazil's total exports and imports in billions of USD from 2010 to 2017. As seen from the graph, the total value of exports and imports fell sharply from 2014, and reached their lowest levels within this period in 2016. After that, I observe a recovery in both exports and imports, starting in 2016.[5] This dip reflects the deep recession that Brazil went through starting in 2014. During this period, GDP shrank by 3.8% in 2015 and unemployment rose from 6.5% in late 2014 to 11.2% in the second quarter of 2016 (ECLAC, 2016). As confirmed by this graph, the trade balance in 2015 became positive with a surplus in the external sector of 17.67 billion USD. This was mainly driven by a sharp drop in imports of 25% and a less severe drop in exports of 15.2%. As a result, the current account deficit dropped from 104.2 billion USD in 2014 to 58.9 billion USD in 2015 (ECLAC, 2016). The backdrop of this recession was the political crisis that resulted from the loss of congressional support for the government and the lawsuits around the corruption scandals that gripped many parties and politicians. In 2016, a new administration came into place, which emphasised tighter restrictions on public expenditure and which also announced plans to decrease the public debt, which had soared during the crisis (with public sector deficit reaching 10.4% of GDP and primary deficit at 1.9% of GDP). After 2016, Brazil experienced a recovery in both exports and imports.

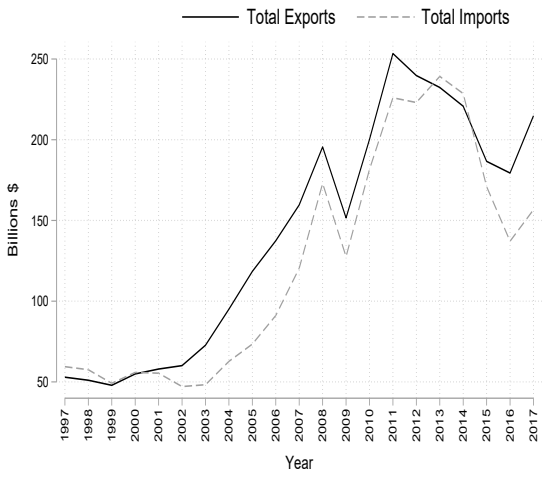
Figure II.3(b) shows the spatial distribution of municipalities registered in my dataset as having at least one importing or exporting firm. There is a concentra-

[5]My empirical strategy accounts for those recession years through year and municipality fixed effects.

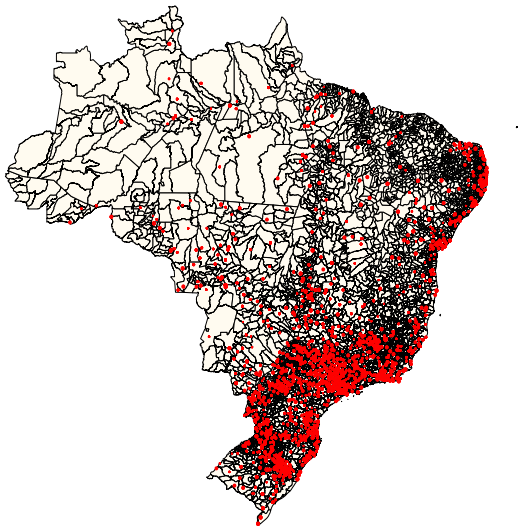
tion of trading firms in the south and south-east region in Brazil. This is where the largest cities such as São Paulo, Minas Gerais, and Rio de Janeiro are located. What is apparent from this graph is the spatial concentration of trading firms, which also implies that there is internal trade within Brazil. The origin of demand for imports may arise from municipalities that are located far from the importing municipalities. The main imports for Brazil during 2010-2017 were petroleum oils from Nigeria, United States, Saudi Arabia, India, Algeria among others, as well as motor vehicles from Argentina and Mexico, hydrocarbon material (can include pesticides) from Bolivia and electronics from China. The main exports for Brazil during this same period were soya beans to China, petroleum oils to the United States, India, Chile and Saint Lucia, and iron ore to Japan, Italy and South Korea.

Figure II.3: Trade in Brazil

(a) Total Imports and Exports to and from Brazil



(b) Spatial Distribution of Trade



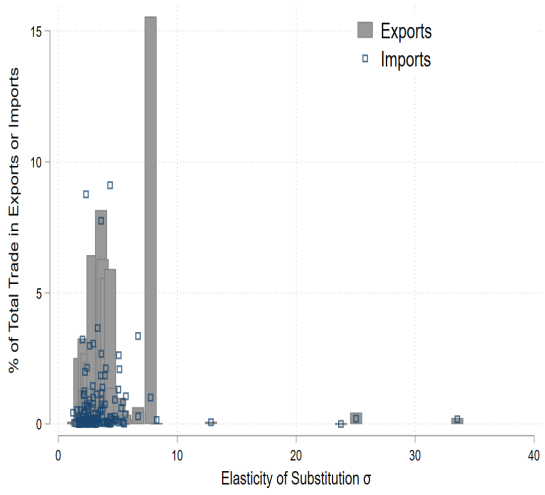
Notes: Figure II.3(a) shows the evolution of total exports and imports in billion of USD from 2010 to 2017. Figure II.3(b) shows the spatial distribution of trade, combining both imports and exports, across Brazil. Every red dot represents a municipality. Author's analysis of the COMEX dataset.

To get an idea on the distribution of Brazilian trade at different levels of differentiation, I plot the share of Brazilian exports and imports by elasticity of substitution σ (Broda and Weinstein, 2006) in figure II.4(a). I notice that there is a large concentration of imports and exports for moderately and less differentiated products (values of σ, the elasticity of substitution of the good, lying between 2 and 10). The graph shows a lot of diffused outliers with the majority of trade happening for levels of σ below 10.

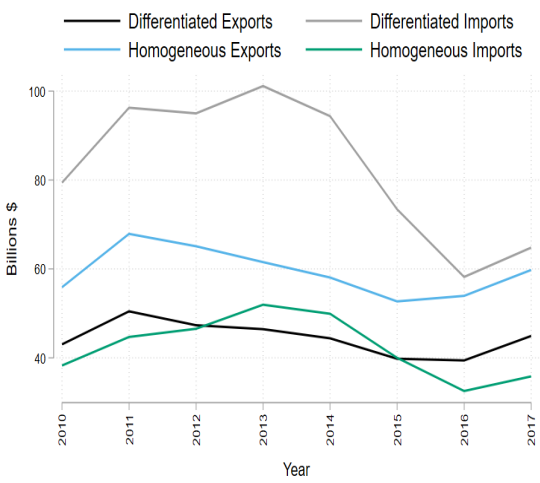
I also plot the evolution over time of differentiated and homogeneous exports and imports, using the classification of Rauch (2009), in figure II.4(b). Consistent with figure II.3(a), I find that all trade faced a sharp reduction with the onset of the recession in late 2014, with a marked reduction in imports (both differentiated and homogeneous).

Figure II.4: Descriptive Graphs for disaggregated Trade Variables by Sector Elasticity of Substitution σ and Rauch Classification

(a) Elasticity of Substitution σ



(b) Differentiated and Homogeneous traded goods



Notes: Figure II.4 calculates the % traded goods by each elasticity of substitution from 2010-2017. The y-axis portrays the percent of each good as a ratio of total for each certain level of σ. Figure II.4(b) depicts (in logs) the evolution of differentiated and homogeneous goods with time based on Rauch (2009) classification of traded goods.

3 Data

My analysis covers the period 2010-2017 and relies on the following datasets. First, the migration data comes from the new database Sistema Nacional de Cadastro e Registro de Estrangeiros (SINCRE). SINCRE registers all foreigners entering Brazil with a visa, except those with tourist visas, and those who apply for the National Migration Registry (RNM).[6] It includes information on country of birth and municipality of destination. However, it is not able to estimate the total population of foreigners

[6]The data used in this paper is based on that accessed on November 2019. This information constitutes the administrative records of the Brazilian Federal Police. Under the framework of the Foreigners' Statute, until 2017, these records were part of the National Registration System for Foreigners' Registration (SINCRE). As of 2018, with the approval of the New Migration Law, they received the name of National Migration Registry System (SISMIGRA). In the database of both information systems are all "immigrants who applied for registration to issue their National Migration Registry (RNM)", formerly the National Registry of Foreigners (RNE) (OBMigra, 2020) (Jannuzzi, 2017).

residing in Brazil as it only captures individuals registered by the police in certain years. The data also doesn't capture asylum seekers and undocumented immigrants who do not possess an RNM. Second, for Brazilian foreign trade statistics, I use COMEX STAT, which contains information on the state and city tax domicile of exporting and importing firms, as well as traded commodities with different countries, classified at the 4-digit level of the Harmonised System.[7] A limitation of the data is that I only observe transactions if a certain good is recorded as a buyer or seller in a municipality. With the absence of information on internal trade, this means that the production of the product or demand for a product may or may not align with the recorded municipality of exports or imports. I describe in section 6 how this limitation affects the interpretation of my analysis. I also miss undeclared cross-border or international flows, but I assume that this data is representative of all trade in Brazil.[8]

I am able to classify traded products as homogeneous and differentiated, using an approach suggested by Rauch (1999). Rauch (1999) shows that an important element in understanding trade relations is studying the type of the traded good in terms of its level of differentiation. He groups goods into three categories: homogeneous, reference and differentiated goods. Homogeneous goods are those goods that are traded at organised exchanges (for example, unwrought lead (SITC 6851) is traded on the London Metal Exchange). Reference goods are those whose prices are listed only in trade publications such as the Chemical Marketing Reporter and include products such as Polymerisation and Copolymerisation Products (SITC 583) or Polyoxyethylene Sorbitan Monostearate. Reference goods are those that are not traded in an organised exchange but have reference prices. These products have a higher degree of differentiation than the homogeneous ones sold on organised exchanges, but are essentially substitutable (Besedes and Prusa, 2006). Since both homogeneous and reference products have a reference price, either on organised exchanges or in trade publications, for the sake of my analysis, I group these two types of products together under one title, homogeneous. Differentiated products do not have a reference price; an example might be 'footwear', which includes a broad range of products that are distinct from the perspective of the consumer. By construction, goods that are neither homogeneous goods nor reference goods, are defined to be differentiated products. My

[7]The Harmonised System is an international nomenclature for the classification of products. At the international level, the Harmonised System (HS) is a 6-digit code system. The first two digits, HS-2, identifies the overall category of the goods, the second two digits, HS-4, identify groupings within those categories, and the last two digits, HS-6, indicate more specific categories (Source: Doing Industry Research: A Resource Guide by the Library of Congress).

[8] after cross-checking data statistics vis a vis other sources such as OEC world.

analysis assigns differentiated products into three broad groups based on their elasticity of substitution σ, following the work of Broda and Weinstein (2006) and Peri and Requena-Silvente (2010), in the following way: highly differentiated when $\sigma \leq 2.1$, moderately differentiated when $2.1 < \sigma \leq 3.5$ and less differentiated when $\sigma > 3.5$.[9]

Finally, I obtained municipal characteristics from the Instituto Brazileiro de Geografia e Estatistica (IBGE) datasets to create a rich set of time-varying controls. The estimations include all of the municipalities in Brazil for which data was available for 2010 to 2017. In Table II.1, I provide the summary statistics for four different samples. The first represents trade with Mercosur countries (those are countries within the same South American Trade Bloc), the second represents trade with countries that belong to the Organisation for Economic Co-operation and Development (OECD), the third represents trade with all other countries and the fourth represents trade with the full sample of countries that have migrants in Brazil.[10] Table II.2 shows the distribution of immigrants in Brazil across municipalities and regions. Since the South-East region is considered to have the richest municipalities in Brazil, it is not surprising to see immigrants concentrated there.[11]

[9]In order to do this, I use the product elasticities for Brazil, at the Harmonised System by Broda and Weinstein (2006), and merge it, using the correspondence tables between the Harmonised System and the standard trade classification, SITC, with the trade data from COMEX. The thresholds based on the product elasticities specified earlier are based on the median values of the elasticity of substitution in table VI in Broda and Weinstein (2006), which are aligned with the Rauch liberal classification of goods.

[10]Mercosur represents the South American Trade bloc that was established in 1991. Its full members are Argentina, Brazil, Paraguay, and Uruguay. Venezuela's full membership was suspended in 2016. Associate members are Bolivia, Chile, Colombia, Ecuador, Guyana, Peru and Suriname. The bloc imposes a common external tariff, a free intra-zone trade and a common trade policy between member countries.

[11]Although Brazil is not a member of the OECD, it is a worthwhile category of countries to include in this descriptive table due to their high level of economic development.

Table II.1: Average Trade Across Municipalities

	Mean Across Years (in Billions of $)			
	Differentiated Exports	Homogeneous Exports	Extensive Exports	Intensive Exports
Country Groupings				
Mercosur	15.601	11.083	2.369	31.233
OECD	21.818	75.401	9.625	115.884
Other	7.056	25.598	7.055	49.749
Total	14.825	37.361	6.350	65.622
	Differentiated Imports	Homogeneous Imports	Extensive Imports	Intensive Imports
Mercosur	7.759	8.960	1.954	23.384
OECD	67.341	35.364	10.657	119.229
Other	7.720	25.790	4.251	35.927
Total	27.607	23.371	5.620	59.513

Notes: This table shows the average of exports and imports across municipalities in Brazil from 2010 -2017, in millions of USD. Th averages are taken over the sample of municipalities that trade a certain category of exports or imports. The extensive and intensive margins are at the product-municipality-country level. Extensive margin represents trade of a certain good for the first time between a municipality and country within 2010-2017 time period. Each of the averages presented in this table refelcts a different set of municipalities. The average of the value of trade for the universe (total) of immigrants is less than other sub-sample country categories because the former are dispersed over a larger number of municipalities and over ones that have low values of trade. Source: Author's analysis of the COMEX data.

Table II.2: Locations of All Migrants Across Brazil by Year

	Percentage				
	North (%)	North East (%)	Center-West (%)	South (%)	South East (%)
year					
2010	7.4%	20.0%	10.6%	24.9%	37.1%
2011	8.1%	20.1%	11.2%	25.0%	35.6%
2012	8.4%	19.8%	10.9%	25.3%	35.5%
2013	9.0%	27.3%	9.3%	23.7%	30.7%
2014	7.4%	22.9%	10.3%	27.6%	31.8%
2015	7.7%	24.3%	9.8%	26.2%	32.0%
2016	8.9%	25.2%	9.2%	25.8%	30.9%
2017	9.1%	24.5%	8.8%	25.6%	31.9%

Notes: This table shows the distribution of immigrants across main regions in Brazil. In particular, the table shows were migrants located as a percentage of total migrants, percentage of border municipalities, and median area and population. Author's analysis of the SINCRE data.

4 Channels of Impact and Empirical Evidence

Migration impacts trade via changes in fixed trade costs and preferences, which may vary depending on the elasticity of substitution of the traded good. My empirical analysis in sections 4.2 to 5.2 study the empirical relationship between trade and immigration for goods at different levels of product substitution, based on Rauch Classification and Broda and Weinstein (2006) elasticity of substitution σ. However, before delving into that, I attempt to estimate empirically the impact of immigration on trade based on the elasticity of substitution of goods. I then link those estimates to the model predictions to visualise the channels of impact based on the degree of product differentiation. In order to do so, I replicate the analysis in Tai (2009) for the Brazilian context. I select this model for the following three reasons. First, the model allows immigrants to impact trade through their influence on preferences, which depends on the degree of product differentiation, σ. Second, the model presents a simple empirical exercise to link the channels of impact from the model to the data, and provides useful predictions that will help rationalise and guide the discussion. And third, the model delivers an equation describing aggregate trade that serves a theoretical underpinning for the main estimation equation in empirical section 4.2.[12]

The model presented by Tai (2009) is a monopolistic competition model with firm heterogeneity and a multi-sector economy. I flesh out the main components of the model in the Appendix II.C.1. Most relevant to the following analysis, the model delivers the following equation that expresses aggregate exports through the following determinants:

$$X_{ij} = \mu_h \times \frac{Y_i Y_j}{Y} \times a_{ij}^{\sigma-1} \times \frac{w_i \tau_{ij}}{\theta_j} \times (f_{ij})^{-(\frac{\lambda}{\sigma-1}-1)} \tag{4.1}$$

Where X_{ij} represents trade between country j and municipality i. Sector specific elasticities of substitution are represented by σ.[13] λ is a sector specific constant, Y_i and Y_j represent GDP of country j and municipality i, a_{ij} is a preference parameter that affects foreign trade between i and j, and w_i represents the wage. Fixed costs are denoted by f_{ij}, and variable costs (iceberg) are denoted by τ_{ij}.

[12]The model and the simple empirical exercise that follow, are useful for the understanding of the channels of impact of trade, and potential interpretation of the results in later section. However, the model, due to its specific structure, is not necessary for the understanding of how immigration affects trade.

[13]This represents the product elasticity of substitution provided by Broda and Weinstein (2006), which is used for the empirical analysis.

4.1 Channels of Impact

The model provides relevant predictions on the channels of impact of immigrants on trade. Migration affects net exports through two channels: preferences and fixed trade costs. Both of these channels are mediated by the degree of substitutability of the traded good. The impact of immigration on trade, as it varies by σ, can be defined as such:

$$F(\sigma) = \frac{d\ln(\mathrm{X}_{ij})}{d\ln(\mathrm{mig}_{ij})} = \underbrace{\frac{\partial\ln(\mathrm{X}_{ij})}{\partial\ln(a_{ij})} \times \frac{\partial\ln(a_{ij})}{\partial\ln(\mathrm{mig}_{ij})}}_{\text{preference component}} + \underbrace{\frac{\partial\ln(\mathrm{X}_{ij})}{\partial\ln(f_{ij})} \times \frac{\partial\ln(f_{ij})}{\partial\ln(\mathrm{mig}_{ij})}}_{\text{fixed cost component}} \tag{4.2}$$

Equation (6.4) can be written as follows (Tai, 2009):

$$F(\sigma) = \tau_a \times \alpha_m + \tau_f \times \theta_m \tag{4.3}$$

where $\tau_a = \frac{\partial\ln(\mathrm{X}_{ij})}{\partial\ln(a_{ij})}$, $\alpha_m = \frac{\partial\ln(a_{ij})}{\partial\ln(\mathrm{mig}_{ij})}$, $\tau_f = \frac{\partial\ln(\mathrm{X}_{ij})}{\partial\ln(f_{ij})}$, and $\theta_m = \frac{\partial\ln(f_{ij})}{\partial\ln(\mathrm{mig}_{ij})}$.

Preferences The effect that the preferences parameter, a_{ij}, has on trade is moderated by the degree of the elasticity of substitution of the good. Goods that have more readily available substitutes will be more impacted by preferences as opposed to those were substitutes are hard to be found.[14] To be precise, the elasticity for preferences is:

$$\frac{d\ln X_{ij}}{d\ln a_{ij}} = \sigma - 1 \tag{4.4}$$

Fixed Costs Trade will increase when fixed trade costs are reduced. The reduction in the latter comes from migrants' ability to bring to the host country specialised knowledge about markets, businesses, and opportunities from their home country as well as provide better contract enforcement.[15] Those main channels through which migration impacts costs, are better characterised as fixed costs to be overcome. This

[14] However if $\sigma = 1$, then the trade channel is not impacted by preferences because the varieties are not substitutable. Note that the minimum value for σ is 1.23 which is strictly greater than 1, therefore the impact of preferences on trade is always positive.

[15] There has been debate into whether costs associated with contract enforcement are fixed or variable costs. In this analysis, contracts are conceptualised as those that are done before a trade exchange happens and is independent from the volume of trade.

increase can occur through the firm intensive and extensive margins.[16] At the intensive margin, this effect is magnified, the larger the elasticity of substitution for the goods: competition is higher when products are more substitutable, a reduction in trade costs will create larger variation in exports. At the extensive margin, new firms will start to export as trade costs are reduced. Firms exporting products with a low σ, higher degree of differentiation, will capture larger shares in the market, as the differentiated product can protect from competition. However, when the firms are trading goods with a high σ, competition is high and hence firms will reach a smaller proportion of the market.

The elasticity for fixed trade costs is:

$$\frac{d \ln X_{ij}}{d \ln f_{ij}} = 1 - \frac{\lambda}{\sigma - 1} \tag{4.5}$$

Channels of Impact in the Brazilian Context Next, I turn my attention to provide a simple exercise that empirically links the aforementioned channels of impact to the data. The purpose of this exercise is to provide insights on how the impact of immigration affects the two channels, fixed trade costs and preferences, at different degrees of the traded good. This is useful because it allows me to hypothesise potential underlying mechanisms through which immigrants affect trade; or the reduced form results that will be presented later in the paper.

I replicate the approach in Tai (2009) in order to assess how migrants impact trade depending on the goods' elasticity of substitution, σ. More specifically, to estimate the results of the disaggregated sector regressions, I pool sector based trade considering each sector as a good with differentiated varieties and interact the respective elasticity of substitution with the migrant stock in each municipality. I estimate the following equation, restricting the power on σ to 3 and including a large set of fixed effects:[17]

$$F(\sigma) = X_{\sigma_{i,j,t}} = \upsilon_{i,t} + \gamma_{j,i} + \phi_{i,t} + f_1 mig_{i,j,t} \times \sigma + f_2 mig_{i,j,t} \times \sigma^2 + f_3 mig_{i,j,t} \times \sigma^3 + \sigma + \sigma^2 + \sigma^3 + \epsilon_i, \tag{4.6}$$

Where $X_{\sigma_{i,j,t}}$, hereafter $X_{\sigma,i,j,t}$ for notational simplicity, is the value of exports or imports varying by the relative sector's elasticity of substitution σ, and the explanatory

[16] The model ignores the impact of immigration on variable trade costs. Since I am not able to show empirical evidence on the extensive and intensive margins of trade at the firm level, due to the type of data I use, I also abstract from this discussion in this analysis.

[17] I tested higher powers and found them to be insignificant.

variables in equation (4.6) are as defined in equation (II.C.37). I use the sample of traded goods, excluding outliers beyond the 95$^{\text{th}}$ percentile, and I find significant impact on total trade (summation of exports and imports). I report the results in the appendix in table II.25, and use the estimates to disentangle the channels as per the model predictions.

The key point estimates are f_1, f_2, and f_3. Therefore, equation (4.6) can expressed as $F(\sigma) = \frac{\partial \ln(Trade)}{\partial \ln(mig)} = 0.372 \times \sigma - 0.161 \times \sigma^2 + 0.019 \times \sigma^3$. I use the point estimates to visually represent of α_m and θ_m as a function of σ.[18,19]

Hypothesis *Immigrants have the highest impact on preferences for highly differentiated goods, while they have the highest reduction on the fixed costs associated with less differentiated goods.*

I plot the preferences and cost channels of impact in figure II.5. I first notice that the channel of impact of migration on preferences decreases with σ. This implies that migration affects preferences in sectors with a higher degree of differentiation (lower σ). What is at play here is the transmission of preferences and information about differentiated products, which makes sense intuitively as those are the goods that are differentiated enough to allow for heterogeneity of preferences. However, the impact of migration on reduction in costs is higher with higher levels of σ, so for more homogeneous goods.[20] In this case, migration is reducing the fixed costs of trade for the goods where the supplier-buyer match is less complicated than say for differentiated goods where trade barriers might be harder to overcome in the latter case. This effect then is driven by the intermediation effect that a migrant can play, and is higher the more straightforward the transaction (a more likely case for homogeneous goods). However, it is worth noting that the impact of the channels is likely to differ

[18] I assume that $\lambda = 9$ and $\beta = 50\%$. This is in order to maintain the condition in the model discussed in Tai (2009) that $\lambda > \sigma - 1$. Assuming that $\beta = 50\%$ represents the case where the two channels are allowed to have an equal impact.

[19] Weighting the two terms of the linear combination in this equation by β then the equation can be written as:

$$F(\sigma) = \beta \times \tau_a \times \alpha_m + (1 - \beta) \times \tau_f \times \theta_m. \tag{4.7}$$

Using equations (4.4) and (4.5), it follows that:

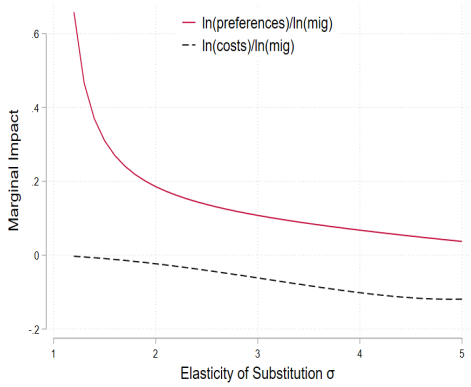
$$\alpha_m = \beta \times \frac{F(\sigma)}{\sigma - 1}, \tag{4.8}$$

$$\theta_m = (1 - \beta) \times \frac{F(\sigma)}{1 - \frac{\lambda}{\sigma - 1}}. \tag{4.9}$$

[20] Note that figure II.5, portrays the impact of immigration of the *reduction* of trade costs.

for imports or exports, with the preferences channel being more dominant for imports than exports.[21]

Figure II.5: Impact of Migration on Trade: Costs and Preferences Channels



Notes: Graph II.5 depicts the impact of migration on preferences and on fixed trade costs as a function of the elasticity of substitution of the traded good σ. The impact of trade is plotted against the elasticity of substitution of the traded good σ up to the value of the 95^{th} percentile. This maintains that I do not divide by zero due to the λ value assumption and also shows the channels of impact for a large majority of trade. Both graphs are based on the model and the empirical estimation.

4.2 Empirical Evidence on the Relationship between Trade and Immigration

There have been many studies, motivated by a gravity model, that estimate the impacts of migration on trade (Zhang, 2020; Bulawayo, 2020; Baltagi et al., 2014; Wang and Badman, 2016; Peri and Requena-Silvente, 2010; Bardhan and Guhathakurta, 2004; Co et al., 2004; Dunlevy, 2006; Millimet and Osang, 2005; Bandyopadhyay et al., 2008; Tadesse and White, 2008; Herander and Saavedra, 2005). I estimate in this section the relationship between trade and immigration using OLS and PPML using a rich and demanding set of fixed effects. This estimation can be thought of as being derived from the augmented gravity equation (4.1) in the model. In particular, equation (4.1) can be fleshed out as follows[22]:

I begin by defining the underpinnings of the equation (4.1) that is relied on mainly in section 4.2, as presented in Tai (2009). In particular, the parameters of preferences

[21]The analysis from this section remains only indicative and provides one potential hypothesis of the underlying mechanisms of the empirical evidence that I will present later on how immigrants affect trade.

[22]The specific derivation from Tai (2009) model is presented in the appendix.

(a_{ijt}), fixed costs (f_{ijt}) and variable costs (τ_{ij}) are captured by:

$$a_{ijt} = mig_{ijt}^{\beta} \times distance_{ij} \times e^{language_{ij}+colony_{ij}+border_{ij}} \tag{4.10}$$

$$f_{ijt} = mig_{ijt}^{(1-\beta)} \times e^{language_{ij}+colony_{ij}} \tag{4.11}$$

$$\tau_{ij} = distance_{ij} \times e^{language_{ij}+colony_{ij}+border_{ij}} \tag{4.12}$$

Building on the aforementioned definitions, equation (4.1) can be re-written as such:

$$lnX_{i,j,t} = \zeta ln(gdp_{it} \times gdp_{jt}) + \beta ln(mig_{ijt}) + \gamma_{ij} + \epsilon_{ijt} \tag{4.13}$$

In order to estimate equation (4.13), I estimate the following equation using a rich set of fixed effects as defined below:

$$X_{i,j,t} = \upsilon_{it} + \gamma_{ij} + \phi_{jt} + \beta ln(mig_{ijt}) + \epsilon_{ijt} \tag{4.14}$$

Where X_{ijt} is the natural logarithm of exports or imports in municipality i to country j. I am interested in identifying, to the extent possible, the point estimate β, and so I employ a rich set of fixed effects to control for the other terms in equation (4.1).[23] In particular, $\upsilon_{i,t}$ are municipality-year fixed effects that absorb time-varying characteristics at the level of a municipality, including average wage $log(w_{it})$, population size and GDP, $log(Y_i)$, per municipality. The term γ_{ji} is the country-municipality fixed effect which capture, between municipality i and country j, bilateral variable trade costs the term $log(\tau_{ij})$, assuming they are constant over time, as well as bilateral time-invariant transport costs that can result from distance, geography, culture, and national and local institutions.[24] The term ϕ_{jt} is a country-year fixed effect which absorbs time varying characteristics at a country of origin level, like improvements in the levels of development and technology affecting a certain country, GDP levels given

[23]The exact definition of the dependent variables used in the estimations is provided in the Appendix in table II.24.

[24]Those variable trade costs are usually associated with transport and tariff costs that can be less susceptible to the influence of immigrants (Peri and Requena-Silvente, 2010).

by $log(Y_jt)$, and all trading country aggregate shocks.[25]

After incorporating the aforementioned fixed effects, what is left from equation (4.13) is the terms pertaining to the preferences $(\sigma_h - 1 \times log(a_{ijt}))$ and the fixed costs $\Big((-\frac{\lambda_h}{\sigma_h - 1} - 1) log(f^h_{ijt})\Big)$. Those are two channels through which immigration could have an impact on trade. I estimate equation (4.14) separately on homogeneous, reference and differentiated goods, as well as based on different levels of the elasticity of substitution, σ.

Therefore, mig_{ijt} measures the stock of migrants from country j in municipality i at time t as a share of the municipal population[26], and ϵ_{ijt} is the error term. With this highly demanding equation, the impact of immigration on trade is identified by the variation within a trading country-municipality over time. However, the results from this analysis are best interpreted as correlations because multiple sources of endogeneity threaten a causal interpretation. For one, immigrants do not locate randomly within Brazil and many will be attracted to areas that promise prospering growth. Another concern is reverse causality where by large multinational companies also attract and hire immigrants to work for them. Therefore, I use the most demanding specification of this estimation in terms of fixed effects and I present the impact of trade at the product extensive margin. In the following sections, I use an instrumental variable approach to ameliorate some of the endogeneity concerns.

I estimate equation 4.14 using an ordinary least squares (OLS) and a psuedo poisson maximum likelihood estimator (PPML). This approach follows recent literature on the estimation of (Anderson and van Wincoop, 2003; Feenstra, 2004; Santos Silva and Tenreyro, 2006). In particular, the PPML is preferred to OLS because it accounts for two characteristics in the data: the presence of actual zeros in trade, and heteroskedasticity in the errors.[27] All estimations presented measure the dependent variable as its

[25] Adding municipality-country fixed effects can be seen as "excessive" and potentially drives the results close to zero (Assuncão and Verdier, 2020). Despite this, I choose to include them in the estimation in order to control for potential unobserved bilateral time-invariant factors such as transport costs, distance, cultural proximity or symbolism of a certain municipality, among others. That said, I explore in the appendix a different set of fixed effects where instead of municipality-country and municipality-year fixed effects, I add state-country, where the state is a higher geographic level. Brazil has 27 states and 5570 municipalities. I report the results in the appendix.

[26] As expected, this variable has a lot of zeros across the municipalities. So not to lose observations when using logs for this term, it is customary to add one to the variable before taking logs. Note however that since the ratio of migrants to the population is a lot less than one, $ln(mig_{ijt} + 1) = mig_{ijt}$. So to ease the interpretation as a one percentage point change, I use the form of this variable in terms of $mig_{ijt} \times 100$. That said, I have confirmed the equivalence of the estimates on all the reported results across the measurements of this immigration variable.

[27] Santos Silva and Tenreyro (2006) show that the multiplicative form of the gravity equation has a stochastic error term that would violate the OLS homoskedasticity conditions. PPML coefficient interpretation is similar to that of the OLS when the dependent variable is in logarithmic form.

logarithm plus one. Standard errors are corrected for heteroskedasticity by White's (1980) method and clustered at the municipality level. In order to account for the fact that the baseline level of trade is now different for areas that had zero immigrants, I include in the estimation equation (4.14) a dummy that indicates whether the municipality had no migrants. The regression is weighted by the municipality population. Therefore, in this estimation per (4.14), the impact of immigration on trade is identified by the variation within a trading country-municipality over time.

4.2.1 Results

I present the estimates of equation 4.14 in table II.3. As noted above, I also include a dummy variable for the presence of any migrants in a municipality and find it to be statistically significant. This indicates that there is a discontinuous change in the impact of migrants on trade when going from municipalities with no immigrants to municipalities with some immigrants. I explore the impact on the extensive margin at the product level. I define this extensive margin at the product level as all products that started being traded to a certain country during the period of analysis (2010-2017) but were not traded prior to 2010. I present the results on the extensive margin in tables II.5 and II.6.

Based on OLS estimations, my estimates based on real USD value of traded goods indicate that there is a positive correlation between immigrants and differentiated imports. I also find a negative correlation between immigrants and homogeneous imports and exports, and between immigrants and differentiated exports.[28] However, these correlations are not significant[29]. Based on the more robust PPML estimations, I find that the presence of immigrants is positively correlated with imports at different levels of differentiation, and with exported products at higher degree of differentiation, albeit the impact is statistically insignificant. But I find that immigration is negatively correlated with exports, but positively correlated with imports, for products at low level

[28] Although my results do not reflect elasticities, it is worthwhile to mention some elasticities found in the literature. For example, Briant, Combes, and Lafourde (2009) estimate an elasticity of exports to migrants of 0.07-0.10 in the French context, while Wagner, Head and Reis (2002) find an elasticity of 0.09 when studying Canadian provinces. Elasticities differ by country studied and method used, however the overall mean of the immigration elasticity of exports is found to be 0.17 while that of imports, 0.16 (Genc, 2014).

[29] A one percentage point increase in migrant to population ratio leads to $100 \times (e^{\beta} - 1)\%$ change in the outcome variable . Note that the estimated coefficients cannot be interpreted as elasticities in a straightforward manner. The reason is because of two characteristics about the data: 1- the explanatory variable is the ratio of immigrants to the population in each municipality in (%). While the outcome variable is of the form $log(trade + 1)$ and for values higher than one this is $\approx log(trade)$.

of differentiation[30]. I do not find evidence that this effect is driven by products at the extensive margin. However, when I study the impact of immigration on trade at the product extensive margin, I find that immigration is positively correlated with exports and imports at high levels of differentiation as well as positively correlated with highly differentiated imports.

The results I obtain on imports, especially on imports at low and high levels of differentiation at the product extensive margin based on σ values, is suggestive that migrants are possibly decreasing fixed costs as well as impacting preferences. Based on the model (implication illustrated in figure II.5), it is possible that the reduction in trade costs channel is more dominant for the increase observed in imports with low differentiation, because the preferences channel might play a much smaller role for such goods. In terms of magnitudes, the impacts on highly differentiated imports is approximately 2.8 times larger than the impact on less differentiated imports. This result could be due to immigrants impacting preferences as well as fixed costs for differentiated imports, which could translate into higher realised impacts on trade, due to the larger share of the market that firms can grasp when trading differentiated versus homogeneous goods (Tai, 2009).

If immigrants can reduce the fixed costs for differentiated exports as well as imports equally, then if the fixed cost channel is prominent, I would expect to see a statistically significant impact on both highly differentiated export and imports. Since the impact of immigration seems to be concentrated on differentiated imports, I hypothesise that if immigrants are impacting more the preferences channel than the fixed costs channel. The negative impact of immigration on homogeneous exports, at the intensive margin, can be explained by immigrants locating in areas where firms export more sophisticated products. Since this analysis is based on a correlation, it is not immune to negative selection bias. Another potential reason, is that the usage of the fixed effect at the municipality-country level can be highly restrictive and reduces variation permitted by the data. As an additional analysis, I re-estimate equation (4.14) using state-country fixed effects as opposed to municipality-country fixed effects. I add a set of municipality-country bilateral controls such as distance and common border dummy. I present the results in the appendix. Based on the PPML estimations, I report a highly significant and positive correlation between immigration and differentiated exports and imports at the extensive margin. Consistently, disaggregating trade based on σ

[30]The mean of the independent variable is 0.0024 (in %). The reason why it is low is because the mean is taken over the universe of municipalities in Brazil that received no immigrants and is biased downwards by having very few immigrants from certain countries and for certain years.

elasticity of substitution, I find a positive correlation between immigration and highly differentiated exports and low differentiated imports. While the impact observed on imports holds when estimating equation (4.13) but varying the level of fixed effects from municipality-country to state-country, the impact on highly substitutable exports doesn't.

Thus far in the paper, I have shown, mainly through correlation analysis, that there exists a positive relationship between immigration and imports, but a weak negative relationship with highly substitutable exports. In the next sections of this chapter, I turn my attention towards attempting to estimate a causal impact of immigration on trade at a municipality and a country-municipality levels.

Table II.3: Effect of Immigration on Trade (in USD) based on Rauch Classification

	(1)	(2)	(3)	(4)	(5)	(6)
	Total Exports	Differentiated Exports	Homogeneous Exports	Total Exports	Differentiated Exports	Homogeneous Exports
OLS and PPML analysis						
Migrants	-0.00374	-0.0189	0.0149	-1.981	0.786	-2.461
	(0.0353)	(0.0295)	(0.0290)	(1.468)	(0.567)	(1.504)
I(Zero Mig)	-0.0121**	-0.00857*	0.000468	-0.0188	-0.0105	-0.0273
	(0.00568)	(0.00498)	(0.00569)	(0.0157)	(0.0236)	(0.0258)
	(1)	(2)	(3)	(4)	(5)	(6)
	Total Imports	Differentiated Imports	Homogeneous Imports	Total Imports	Differentiated Imports	Homogeneous Imports
Migrants	0.0147	0.00797	0.00790	0.895	1.849	0.550
	(0.0140)	(0.0118)	(0.0242)	(0.910)	(1.176)	(0.464)
I(Zero Mig)	-0.0233***	-0.0252***	-0.00383	-0.0333	-0.0583**	-0.0118
	(0.00582)	(0.00567)	(0.00495)	(0.0211)	(0.0238)	(0.0308)
Observations	809344	809344	809344	175032	131833	94692
Sample	All	All	All	All	All	All
Municipalities	4341	4341	4341	4341	4341	4341
Estimation	OLS	OLS	OLS	PPML	PPML	PPML
R squared	0.9670	0.9654	0.9661	0.9677	0.9698	0.9611
Municipality-Year F.E	Yes	Yes	Yes	Yes	Yes	Yes
Country-Municipality F.E	Yes	Yes	Yes	Yes	Yes	Yes
Country-Year F.E	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table shows the OLS and PPML estimates from equation (4.14). The dependent variable is of the form $log(x+1)$ and the explanatory variable of immigration is the ratio of immigrants to the population per municipality (in %). I(Zero Mun) is a dummy variable that indicates 1 if a municipality has no immigration. All estimations include a dummy variable indicating 1 for areas that have zero trade . All estimations include municipality-year, trading country-municipality, and trading country-year fixed effects. The R^2 in the PPML estimations are Psuedo R^2, while those in the OLS estimations are the adjusted R^2. Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1.

Table II.4: Effect of Immigration on Trade (in USD) based on σ Elasticity of Substitution

	(1) log(Exports Low Diff.)	(2) log(Exports Mid Diff.)	(3) log(Exports High Diff.)	(4) Exports Low Diff.	(5) Exports Mid Diff.	(6) Exports High Diff.
OLS and PPML analysis						
Migrants	-0.0182	-0.0201	-0.00771	-2.757*	-0.0407	0.0911
	(0.0380)	(0.0526)	(0.0177)	(1.446)	(0.918)	(0.293)
I(Zero Mig)	-0.00857	-0.00494	-0.00468	-0.0186	0.00747	-0.0118
	(0.00579)	(0.00546)	(0.00476)	(0.0244)	(0.0208)	(0.0790)
	(1) log(Import Low Diff.)	(2) log(Import Mid Diff.)	(3) log(Import High Diff.)	(4) Import Low Diff.	(5) Import Mid Diff.	(6) Import High Diff.
Migrants	0.0280	0.0181	-0.0133	2.220*	0.885	0.383
	(0.0190)	(0.0158)	(0.0317)	(1.218)	(0.754)	(0.382)
I(Zero Mig)	-0.0195***	-0.0191***	-0.00976**	0.00550	-0.0520*	0.00354
	(0.00554)	(0.00564)	(0.00397)	(0.0216)	(0.0302)	(0.0201)
Observations	809344	809344	809344	127369	126802	52602
Sample	All	All	All	All	All	All
Municipalities	4341	4341	4341	4341	4341	4341
Estimation	OLS	OLS	OLS	PPML	PPML	PPML
R squared	0.9648	0.9639	0.9618	0.9708	0.9565	0.9482
Municipality-Year F.E	Yes	Yes	Yes	Yes	Yes	Yes
Country-Municipality F.E	Yes	Yes	Yes	Yes	Yes	Yes
Country-Year F.E	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table shows the OLS and PPML estimates from equation (4.14). The dependent variables in this table represents goods that are traded for the first time from a certain municipality to a certain country between 2010 - 2017. The dependent variable is of the form $log(x+1)$ and the explanatory variable of immigration is the ratio of immigrants to the population per municipality (in %). I(Zero Mun) is a dummy variable that indicates 1 if a municipality has no immigration. All estimations include a dummy variable indicating 1 for areas that have zero trade . All estimations include municipality-year, trading country-municipality, and trading country-year fixed effects. The R^2 in the PPML estimations are Psuedo R^2, while those in the OLS estimations are the adjusted R^2. Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1.

Table II.5: **Extensive Margin:** Effect of Immigration on Trade (in USD) based on Rauch Classification

	(1) Total Exports	(2) Differentiated Exports	(3) Homogeneous Exports	(4) Total Exports	(5) Differentiated Exports	(6) Homogeneous Exports
OLS and PPML analysis						
Migrants	0.00409	-0.0195	0.0111	2.675	1.370	10.90*
	(0.0226)	(0.0213)	(0.0166)	(2.382)	(1.780)	(5.690)
I(Zero Mig)	-0.00110	-0.00549	-0.00495	-0.0467	0.0125	0.215
	(0.00745)	(0.00611)	(0.00574)	(0.0843)	(0.107)	(0.162)
	(1) Total Imports	(2) Differentiated Imports	(3) Homogeneous Imports	(4) Total Imports	(5) Differentiated Imports	(6) Homogeneous Imports
Migrants	0.0449*	0.0404*	0.00626	2.295*	4.010*	-0.0961
	(0.0243)	(0.0212)	(0.0250)	(1.306)	(2.050)	(1.836)
I(Zero Mig)	-0.0358***	-0.0250***	-0.0100*	-0.242***	-0.237***	-0.168*
	(0.00683)	(0.00635)	(0.00567)	(0.0656)	(0.0862)	(0.0893)
Observations	809344	809344	809344	111203	85649	42224
Sample	All	All	All	All	All	All
Municipalities	4341	4341	4341	4341	4341	4341
Estimation	OLS	OLS	OLS	PPML	PPML	PPML
R squared	0.9322	0.9297	0.9162	0.7988	0.8165	0.7981
Municipality-Year F.E	Yes	Yes	Yes	Yes	Yes	Yes
Country-Municipality F.E	Yes	Yes	Yes	Yes	Yes	Yes
Country-Year F.E	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table shows the OLS and PPML estimates from equation (4.14). The dependent variables in this table represents goods that are traded for the first time from a certain municipality to a certain country between 2010 - 2017. The dependent variable is of the form $log(x+1)$ and the explanatory variable of immigration is the ratio of immigrants to the population per municipality (in %). I(Zero Mun) is a dummy variable that indicates 1 if a municipality has no immigration. All estimations include a dummy variable indicating 1 for areas that have zero trade . All estimations include municipality-year, trading country-municipality, and trading country-year fixed effects. The R^2 in the PPML estimations are Psuedo R^2, while those in the OLS estimations are the adjusted R^2. Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1.

Table II.6: **Extensive Margin:** Effect of Immigration on Trade (in USD) based on σ Elasticity of Substitution

	(1)	(2)	(3)	(4)	(5)	(6)
	Exports Low Diff.	Exports Mid Diff.	Exports High Diff.	Exports Low Diff.	Exports Mid Diff.	Exports High Diff.
OLS and PPML analysis						
Migrants	-0.000649	-0.0226	0.000702	5.470*	0.579	-3.158
	(0.0213)	(0.0442)	(0.0122)	(3.132)	(1.749)	(2.303)
I(Zero Mig)	-0.00619	0.00481	-0.00516	-0.126	0.0480	-0.305
	(0.00654)	(0.00606)	(0.00404)	(0.121)	(0.116)	(0.187)
	(1)	(2)	(3)	(4)	(5)	(6)
	Import Low Diff.	Import Mid Diff.	Import High Diff.	Import Low Diff.	Import Mid Diff.	Import High Diff.
Migrants	0.0409**	0.0182	0.0213	3.084*	1.845	8.808**
	(0.0208)	(0.0168)	(0.0223)	(1.583)	(1.286)	(3.762)
I(Zero Mig)	-0.0232***	-0.0167**	-0.00823*	-0.186**	-0.266***	-0.202*
	(0.00653)	(0.00666)	(0.00431)	(0.0890)	(0.0857)	(0.120)
Observations	809344	809344	809344	74835	73524	21116
Sample	All	All	All	All	All	All
Municipalities	4341	4341	4341	4341	4341	4341
Estimation	OLS	OLS	OLS	PPML	PPML	PPML
R squared	0.9228	0.9216	0.8962	0.8242	0.8001	0.8141
Municipality-Year F.E	Yes	Yes	Yes	Yes	Yes	Yes
Country-Municipality F.E	Yes	Yes	Yes	Yes	Yes	Yes
Country-Year F.E	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table shows the OLS and PPML estimates from equation (4.14). The dependent variables in this table represents goods that are traded for the first time from a certain municipality to a certain country between 2010 - 2017. The dependent variable is of the form $log(x + 1)$ and the explanatory variable of immigration is the ratio of immigrants to the population per municipality (in %). I(Zero Mun) is a dummy variable that indicates 1 if a municipality has no immigration. All estimations include a dummy variable indicating 1 for areas that have zero trade . All estimations include municipality-year, trading country-municipality, and trading country-year fixed effects. The R^2 in the PPML estimations are Psuedo R^2, while those in the OLS estimations are the adjusted R^2. Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1.

5 Estimating a Causal Impact

Estimating equation 4.14 make use of a demanding specification that controls for a municipality-country, municipality-time and country-time fixed effects. However, this specification is not immune to potential endogeneity issues pertaining to the non-random location of immigrants. Unobserved factors that are not accounted for by the estimation can be affecting both migration and trade (Hanson, 2010; Lucas, 2006). For example, it can still be the case that the yearly flows of migration are highly correlated with trade flows, or that the location of immigrants is correlated with the location of firms that are productive enough to be able to trade[31]. This potential reverse causality prevents me from interpreting the OLS or the PPML estimates from equation II.C.37 causally.Further, it could be the case that more educated immigrants locate near firms that import or export products with higher degrees of differentiation (which are perhaps associated with a more complex production process). Alternatively, such firms might locate in cities that would allow them to benefit from agglomeration and network effects as well as local amenities that are important considerations for immigrants' locations. In short, the location of immigrants is potentially endogenous and non-random. To ameliorate this concern, I use an instrumental variable strategy in order to estimate a causal impact of immigration on trade (Bratti et al., 2012; Peri and Requena-Silvente, 2010).

In this analysis, I approximate the estimation of the trade equation (4.1). I start in this section by aggregating at the municipality level using equation (5.15), and then, I employ this IV strategy at the country-municipality level and estimate heterogeneous impacts of immigration on trade depending on the country of origin of the immigrant, in section 5.2.

The estimation at the municipality level takes the following form:

$$X_{i,t} = \upsilon_t + \gamma_i + mig_{i,t} + \nu\chi_{i,t} + \epsilon_{i,t} \tag{5.15}$$

In this specification, $X_{i,t}$ is a measure of exports or imports in municipality i at time t in the form of $log(X_{i,t}+1)$, υ_t is year fixed effects, γ_i is the municipality fixed effect, $mig_{i,t}$ measures the number of migrants as a share of the population in municipality i at time t, $\chi_{i,t}$ is a vector of municipality controls. $\epsilon_{i,t}$ is the error term. Standard

[31]The idea that there exists a productivity cut-off that trading firms meet before being able to export is based on Melitz (2003).

errors are corrected for heteroskedasticity by White's (1980) method and clustered at the microregion level.

5.1 Identification Strategy

A standard approach in the migration literature to dealing with the endogeneity of migration flows is to use a shift-share type of instrumental variable strategy. In this case, the shares are constructed using information on the historical location of migrants from the same country of origin and the shifts refer to the new migrants' inflows from these countries (Jaeger et al., 2018). The necessary assumption is that the factors that determined the historical location of migrants do not affect current dynamics.

Migration to Brazil has occurred in waves and there was a very small population of many of the immigrant communities before the mid-2000s. In table II.22 in the appendix, I show for the top 30 countries that arrived to Brazil, the growth in migrant numbers relative to pre-2000 levels. I notice that among these top countries of origin of immigrants, there has been a very significant increase in the number of migrants when compared to their pre-2000 numbers, due to the fact that the absolute numbers of the latter are very small.[32] In the Brazilian context, there has also been new waves of migrants from countries where there were historically no or very few migrants present in Brazil before the year 2000.[33]

This setting is both a gift and a challenge. It is a gift because the large increases in migration and the presence of new migration waves makes Brazil a very interesting country in which to study the implication of immigration. But, it is a challenge because I cannot easily follow the literature in using the historical location of immigrants as my instrument for the destination of new migrants. The low initial numbers of migrants, and the subsequent expansion of migration, including from new countries, together require a different approach. In the appendix (in table II.21), I create the traditional IV shift share instrument based on earlier settlements of migrants sharing the same country of origin. As can be seen from the table, the traditional shift share instrument is a weak instrument in this context due to low spatial variation that some countries had

[32]Some countries stand out with an increase ranging between 312% to 8667% and with the largest increase characterising countries such as Cuba (4,257%), Senegal (8,667%), India (7,063%), and The Philippines (8,218%). Immigrants from Haiti show the largest increase of more than 180,000%. The large increase of immigrant flows from Haiti in recent years was triggered by an Earthquake that shook the country in late 2010, leading to a massive out-migration.

[33]For example, the following countries: Afghanistan, Albania, Benin, Estonia, Jamaica, Congo, Georgia, Maldives, among others, had less than 10 immigrants in Brazil prior to the year 2000, but had more than 100 immigrants arriving into the country between 2000 and 2010.

before 2000. The standard IV approach works best when there are stocks of migrants spread across the country with some spatial dispersion, with subsequent flows that map onto the pre-existing stocks by country of origin. In my context, by contrast, the initial stocks are so low for most countries that they are necessarily not very spatially dispersed. Some countries are represented in the pre-2000 data by only a few immigrants, leaving little or no scope for dispersion across municipalities.

I use a modified version of the shift-share instrument which relies on the linguistic commonality of new migrants with previously settled migrants. This effectively broadens the set of locations that might be attractive to a new migrant. The traditional shift share IV that relies on the commonality of the country of origin assumes that migrant location decisions are influenced by some combination of language, culture, and social networks. My instrument assumes and shows that shared language is a sufficiently strong connection to influence location choices. I use all the migrant settlements that entered into Brazil prior to 2000, 10 years before the period of my analysis. The idea is that having more individuals with a common language in a location can affect the location decisions of new migrants. For instance, French speakers could be likely to settle in locations with a larger number of French speakers, not just people from their country of origin (France, for example). In that sense, I replace country of origin as the common factor across migrants with language similarity based on the majority language spoken in the country of origin.[34] This method is similar in concept to the traditional past-settlement instrument that relies on the historic distribution of the country of origin. Through the commonality of language across countries, I am able to have more historic variation across municipalities that relates to current migrant locations. For example, a person speaking French, will now relate to the location of 12 other French speaking countries as opposed to one, which is the country of origin of that migrant, and a person speaking Spanish will relate to 18 other Spanish speaking countries that located in Brazil pre-2000. In cases where the majority language is spoken by only one country, for example Wolof spoken mainly in Senegal, then in this case, the historic settlements of the country of origin predict the location of immigrants from that country. In this analysis, each country is assigned the language spoken by

[34]It is worth to note that individuals who immigrant may typically speak more than one language, especially the language of the country of the destination and a lingua franca, like English. In this analysis, I rely on the majority language spoken in the country of origin and the intent is to use language commonality as a proxy of culture that transcends the country of origin. My results and first stage F-statistics are robust to the inclusion or exclusion of Portuguese and English immigrant speakers.

the majority of the citizens based on the following sources: World Atlas Language Structures, the CIA The World Factbook and official country by country sources.[35]

As discussed earlier, grouping countries into the language bins gives me more spatial historic variation that gives me more power in the first stage IV estimations. However, it can be the case that the variation and power is coming from a small subset of common languages. I follow the recommendations by Goldsmith-Pinkham et al (2019) to uncovering those language shares with the highest variation and weighting to the IV. I find that 24 of the language groups (or 27% of total) combine multiple countries and hence provide the largest spatial variation. I elaborate on this test and its implications in section II.A in the appendix.

In particular, I create a municipality share for each language ($L_{i,f}$) as such:

$$L_{i,f} = (\frac{Mig_{i,f}}{\sum_f Mig_{i,f}})_{t=0} \tag{5.16}$$

In this case $Mig_{f,i}$ is the number of speakers of language f (e.g. French) in location i at time 0 (i.e. year 2000). In order to get our instrument $Z_{i,t}$ for each municipality, I sum:

$$Z_{i,t} = \sum_{f=1}^{F} L_{i,f} \times \Delta Mig_{f,t} \tag{5.17}$$

Where $\Delta Mig_{f,t}$ is the annual inflow of migrants from a country with the main language spoken f in a given year t. Table II.7 reports the first stage estimations. In both cases the instrument has a strong predictive power over the location of migrant inflows. Table II.7 reports the first stage estimations. I find that the first stage is significant at the 1% level indicating the instrument's strong predictive power across multiple specifications that include different levels of fixed effects (at the microregion, mesoregion, mesoregion interacted with year, and municipality). I find that the Kleiberg-Paap F-statistic is high, and ranges between 12.14 to 30.07, the latter when using municipality

[35]There remains multiple limitations for this approach and I expand on that in section 6. One of those concerns is that assigning for each country, the language being spoken by its majority citizens may not be an accurate representation of the language spoken by the average immigrant from this country. I am unable to account for this in more detail as I do not have data on the migrant's regional location within their country of origin. Future work on this variation of the shift-share instrumental variable strategy should take this into consideration, with one idea using indices of language proximity measures that include the Linguistic Levenshtein distance Automated Similarity Judgment Program (ASJP) or Melitz and Toubal (2012) indices in order to create a weighted average of the different languages spoken. In a separate chapter of this thesis, I experiment with the linguistic proximity index. More generally, the role of culture can be considered more robustly as a potential factor in the location decision of immigrants, but this analysis goes beyond the scope of this paper.

fixed effects. As I am interested in within municipality changes, I employ specification (6) to estimate the second stage results in the subsequent section 5.2 where I use municipality and year fixed effects as well as time varying municipality level control variables.

This specification is consistent with the ones used in the literature and the classical papers on this methodology such as Card (2001) and Altonji and Card (1991). My identification strategy relies on the commonality of language between earlier migrant settlements[36] and new immigrants between 2010 to 2017. The validity of the instrument is based on the assumption that the 2000 migration shares varied across municipalities because of factors related to the persistent agglomeration of migrants in some municipalities. These historical differences, after controlling for a rich set of municipality characteristics, affected the change in recent migration waves but were unrelated to shocks affecting municipality level firm growth, productivity and ability to trade. However, it remains possible that historical immigration sorted into municipalities that had different economic trends unrelated to the recent immigration waves, or that such municipalities had persistent local characteristics that are correlated with new incoming immigration waves and economic factors impacting trade.

Hence, the exclusion restriction relies on the exogeneity of the historic shares of immigrants' locations. Although I cannot formally test for this exclusion restriction, I follow recommendations by the most recent literature on this matter (Goldsmith-Pinkham et al., 2019) to address empirically some of the concerns surrounding potential violations of this assumption. I begin by unpacking the language groups that are responsible for the highest variation in the calculated instrumental variable. I find that Spanish, Italian and Japanese speakers are the language groups that contribute the most to the variation in the shift share IV, with Spanish speakers accounting for 81% of the positive weight of the estimator. Because there is one language share that has a high weight, I test whether those historic shares are correlated with potential baseline (2000) covariates that would impact trade in recent years and I present the results in table II.16. I find that these main historic language groups are not correlated with GDP levels, tax, employment, value added in agriculture, industry and services. I only find that the Spanish language group is weakly correlated (at the 10% level) with value added in services. However the R^2 is very small and doesn't exceed 0.26,

[36]The specification of the shift-share instrumental variable is very similar to that constructed by Mayda et al. (2022). The authors create shares for both high skilled and low skilled immigrants based on the past settlement instrument that relies on the commonality of the country of origin. Instead, I rely on the commonality of the language spoken by the majority of the population.

which tells me that the historic presence of these language groups explains very little (and mostly insignificant) key baseline covariates.

Furthermore, I also test and find that there is no correlation between immigrants' location decisions after 2010 with changes in economic outcomes in the previous decade (between 2010 and 2000) such as GDP, value added in services, industry, agriculture and trade outcomes in 2010. I report the results in table II.17 in the Appendix section II.A. One might still be worried that the historic migrant settlements are correlated with future trade outcomes because of the networks, knowledge and complementarity between labour and trade. In order to address this concern, I test if the historic migrant shares are correlated with pre-trends in trade outcomes for the years leading up to the period I study, and one that represents major increases in immigration. I report the results in table II.18 and find that there is no correlation between historical immigration shares and trade outcomes such as total imports and exports.[37] In the Appendix section II.A, I discuss in detail the analysis based on recommendations by Goldsmith-Pinkham et al (2019). Finally, in my estimation, I add a rich set of time varying economic controls and use fixed effects at the level of the municipality in order to address concerns around persisting endogeneity of immigrants' locations.[38] This level of fixed effect will allow comparison to be within the same municipality.

Importantly, my results study the impact of immigration at the product extensive margin. If there is impact at this margin, then I have a stronger case that immigrants are contributing to trade through the generation of new products and helps reduce the endogeneity concern of reverse causality. In addition, I study a specific group of countries who had a very meagre presence, less than 15 immigrants, before 2000, but had much larger immigration flows in recent years. This exercise serves two purposes. First, it helps reduce concerns of reverse causality that can happen when foreign owned or multinational corporations hire immigrants into the country. For example, one might be concerned that a German company established in Brazil by early German settlements is hiring German immigrants. This would violate the exclusion restriction.

[37] I also test for other trade outcomes disaggregated by the type of product and I find consistent results that historic migrant shares are uncorrelated with trade pre-trends.

[38] I also estimate the results varying the level of fixed effect, to administrative levels higher, such as the microregion and mesoregion. I find the results on the impact of immigration on trade to hold in these specifications as well.

Table II.7: First Stage Estimates

	(1) Migrants	(2) Migrants	(3) Migrants	(4) Migrants	(5) Migrants	(6) Migrants
IV First Stage						
IV	0.00530***	0.00581***	0.00564***	0.0157***	0.0155***	0.0134***
	(0.00148)	(0.00161)	(0.00162)	(0.00287)	(0.00291)	(0.00272)
Observations	44543	44543	44535	44560	44543	44543
Kleibergen-Paap F-stat	12.7284	13.0455	12.1498	30.0761	28.4331	24.1231
Cragg-Donald F Stat	205.1108	264.7464	249.0449	354.0787	338.4560	215.8885
Municipality Controls	Yes	Yes	Yes	No	Yes	Yes
Year FE	Yes	Yes	No	No	No	Yes
Fixed Effect	Microregion	Mesoregion	Mesoregion X Year	Municipality	Municipality	Municipality

Notes: The table shows the 2SLS first stage estimates based on the estimation of equation (5.15). Robust standard errors in parentheses clustered at the mesoregion level. Regressions include mesoregion fixed effects and time varying municipality controls. Mean of migrant represents the mean of the ratio of the migrants to the population across all municipalities for all years under study. Instrumental variable is divided by 1,000 to represent better first stage estimates. First stage Kleiberg-Paap S statistic and second stage estimates are invariant to this division. Data sources: IBGE data for control variables, and SINCRE for migration data. Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1.

5.2 Results

This section provides the main results from the IV empirical estimation at the municipality level. I present evidence on the impact on overall trade, and on traded goods, including those traded at the extensive margin, as classified by Rauch (2009), and on the sector elasticities of substitution provided by Broda and Weinstein (2006). I also discuss how the results differ when performing the analysis for three different country groups: Mercosur, OECD and all the rest. I use those results to highlight the potential channels at play.

Universe of immigrants: I report the results on aggregated imports and exports in table II.8. I notice that the OLS estimates are smaller in magnitude than the estimates of IV 2SLS. This could reflect a downward bias in the OLS estimates – for example, because of negative selection by immigrants which would occur if immigrants moved to areas that would have had lower long-run economic growth, counter-factually. This difference in magnitude between OLS and IV might be reflecting heterogeneous sub-populations and hence the local average treatment effect (LATE) estimated by the IV 2SLS is larger than the average treatment effect (ATE) estimated by OLS, where OLS estimates are biased downwards. The ATE estimate could be driven by omitted variable bias where the location and concentration of immigrants is negatively correlated with omitted variables. The LATE estimate is an average effect among

compliers which in this study are municipalities whose immigrant population located near historical immigrant locations. It is hence possible that those immigrants are different in unobservable characteristics from others who did not choose to locate within Brazil based on historical immigration communities. To get a sense on the negative selection bias as a potential explanation, I compare the more developed regions, South and South East, with the less developed, North, North East and Centre Regions. The more developed regions have a higher coefficient of variation in their GDP per capita, and therefore more opportunities for the presence of negative selection, so if the latter is a potential mechanism, I would expect the OLS coefficient for the South and South East regions to be more biased than that for the North and Centre regions. I re-estimate equation 5.15, subdividing the samples by regions and find that the OLS coefficient for the North and Centre regions is three times less and more biased downwards than that for the South and South East regions, indicating that negative selection bias is a potential explanation.

Another potential reason why the OLS estimates might be biased towards zero could be as a result of attenuation bias stemming from measurement error in the independent variable, ratio of migrants to population. Also, the instrument might be shifting the behaviour of a subgroup of migrants for whom their impact on trade relationship is higher than average. I am able to reject the null hypothesis of under-identification with the Kleibergen-Paap (2006) *rk* LM statistic. The statistic tests the null hypothesis that the rank of the matrix of the first stage coefficients is 1 against the alternative hypothesis which states that the rank is equal to 2. The estimates reported represent the increase in trade due to a one percentage point increase in migrant share to the population. As noted above, the latter would represent an unrealistically large increase, so I report a more realistic increase that is equivalent to doubling the mean of the migrant share to the population.[39]

The 2SLS IV results suggest that the presence of migrants has a positive and statistically significant impact on differentiated imports. In particular, increasing the migrant share of the population by 0.00062, which would mean doubling the average value of this variable across municipalities, leads to an increase in differentiated imports by 0.022%. Table II.11 provides the IV 2SLS results for traded goods, divided

[39]This is interpreted as an increase of the migrant share by its average value. I take this to be equivalent to doubling the average value of this share of migrants of the population across municipalities, and is calculated as such: $100 \times (e^{\beta \times (\overline{\frac{migrants}{pop}})} - 1)$.

by Rauch (2009) into differentiated and homogenous products at the product extensive margin. I find no evidence that the result on differentiated imports is driven by products at the extensive margin, but is instead driven by an increase in total and differentiated imports at the intensive margin. I also present the results of the impact of immigration on trade, disaggregating the latter based on the elasticity of substitution of the product σ. I find that immigrants have a positive impact on imports with low levels of differentiation at the intensive margin, but not on the extensive margin.

I run a series of robustness checks to test the sensitivity of these results to the sample of municipalities and the level of fixed effects used. I start by using joint Federation Unit (UF) and year fixed effects. I report strong positive impacts due a one percentage point increase on both differentiated and homogeneous exports and imports. I also show that the impact of immigration is stronger for imports at all levels of product differentiation, than that on exports. This is true in terms of statistical significance and point estimate magnitude. When employing fixed effects at the mesoregion, I find consistent results with those found when using fixed effects at the UF level. Further, I estimate equation (5.15), but I exclude municipalities that trade at a level higher than the 95^{th} percentile of the sample at each category of product differentiation, and in another specification, I exclude municipalities at the border. In the sample with excluded outliers, I find no statistically significant impact of immigration on trade, although the positive impact on differentiated imports is qualitatively similar (positive and indicative of a 0.013% increase). This result is suggestive that the significant impact on differentiated imports is driven by municipalities at the upper tail of the trade distribution. Upon studying the impacts of immigration on trade on the sample of municipalities excluding ones at the border, I find that there is a significant and positive increase in differentiated imports, whereby doubling the ratio of immigrant to the population across municipalities, increases differentiated imports by 0.038%. Upon studying the impacts of immigration on trade including only the sample of municipalities that trade within the period 2010 to 2017, I find that the results hold qualitatively, but are not statistically significant. Finally, I estimate equation (5.15) but I interact the value of the control variable in 2010 with a time trend, and find that the positive and significant impact that immigration has on differentiated imports also holds under this specification..

Mercosur and OECD immigrants: I estimate equation 5.15 using IV 2SLS but subdividing the sample into three sets of countries: those belonging to Mercosur,

those belonging to OECD, and those belonging to other countries that aren't in either group. I study the impact of immigrants from those sets of countries, on trade with countries within that specific group. I find that Mercosur immigrants cause a reduction in differentiated imports and I find that this impact holds at the product extensive margin. An equivalent to doubling the average of Mercosur immigrants leads to a 0.019% reduction in differentiated imports. The impacts on trade with OECD from immigrants within that trade block are different to those observed with Mercosur immigrants. Immigrants from OECD countries cause an increase in total imports, as well as imports at medium levels of differentiation. I also find that these impacts hold at the product extensive margin. Immigrants from other countries than Mercosur and OECD, cause an increase in less differentiated imports, however, this impact seems to be driven the intensive rather than the extensive product margin.

Immigrants from new countries of origin: I also study the impacts on trade due to the arrival of immigrants from new countries of origin that had meagre to no presence historically, but had large inflows into Brazil in recent years (2010-2017). I find a positive and significant impact of immigration on homogeneous exports as well as more differentiated exports and imports. However, these result do not hold for products traded at the extensive margin. I report the results in section II.C.6 in the Appendix. Finding consistent results, especially on differentiated imports, using sample of countries, helps ameliorate one of the endogeneity concerns that the IV strategy helps overcome: reverse causality, where historically immigrants established trade relationships that remain persistent in recent years and drive more immigrant inflows.

Linking immigrants' country of origin to trading country: I estimate equation (5.15) at the country, municipality and year levels. This level of estimation allows me to link immigrants in each municipality to trade from and to their country of origin. In particular, I estimate the following equation, which is more in line with the estimation in section 4.2.[40]

$$X_{i,j,t} = \upsilon_{i,t} + \gamma_{i,j} + \delta mig_{i,j,t} + \epsilon_{i,j,t}. \tag{5.18}$$

[40]This level of estimation is similar to the correlation analysis discussed in section 4.2, however I do not use country-time fixed effect in this estimation because I lose a lot of power in the first stage and the Kleiberg and Paap F statistic falls significantly below 10.

In equation 5.15, j represents each country. $X_{i,j,t}$ is a measure of exports or imports between municipality i and the countries j, $\upsilon_{i,t}$ is municipality and year fixed effects, $\gamma_{i,j}$ is the municipality and country fixed effect, $mig_{i,j,t}$ measures the number of migrants as a share of the population belonging to the country j, and $\epsilon_{i,j,t}$ is the error term.

Next, I adjust the IV shift share to match the form of the main explanatory variable in equation 5.21 as follows:

$$L_{i,f} = (\frac{Mig_{i,f}}{\sum_f Mig_{i,f}})_{t=0} \tag{5.19}$$

$$Z_{i,f,t} = L_{i,f} \times \Delta Mig_{f,t} \tag{5.20}$$

In the latter specification, f is the language spoken in country j. So in this case, one language group historically located in Brazil can predict the location of immigrants arriving recently from any country j that has majority population speaking language f. The instrumental variable strategy is similar to the one in equation 5.5. However, the instrument is different in the sense that now for every country, the instrument becomes the multiplication of this share $L_{f,i}$, which represents the ratio of migrants speaking language f in municipality i to total immigrants in i in 2000, by the migrants belonging to that same country group and language in every year $\Delta Mig_{f,t}$. Therefore, what predicts an incoming immigrants location is the historic distribution of co-lingual immigrants interacted with the total yearly inflow of co-lingual immigrants. This instrument varies by the language spoken by the majority of the population in the country of origin of the migrant. The validity of this instrument rests on the assumptions discussed and tested in section 5.2, and the exclusion restriction remains the same: that historic immigration settlements are correlated with current migrant locations, but have no correlation to factors influencing trade 15 years later.

I find results that are consistent with what has been found at the municipality level analysis. In particular, I find a significant increase in imports at low and mid levels of differentiation, due to an increase in the ratio of immigration to the population. Interestingly, I also find a significant increase on homogeneous imports based on Rauch classification of goods. In both of these cases, the results also hold at the product extensive margin. In addition, there is evidence that immigration increased total exports at high levels of differentiation, but this result is only driven by products traded at the intensive margin. I report the results in section II.C.3 in the Appendix.

Potential channels: The results on the different classifications of the traded goods can be indicative of the underlying channels through which immigrants are impacting trade. My results for the universe of immigrants, OECD immigrants and newly arriving immigrants point to a significant increase on differentiated imports due to the presence of immigration. When linking immigrants' country of origin to the trading country, I also find that the increase in differentiated imports exists at the extensive margin due to the presence of immigration. This result can be explained through the existence of both channels, the role of immigrants in altering or diffusing preferences in the host community, and in overcoming fixed trade costs. My theoretical framework suggests that preferences can play a big role for goods that have a high degree of differentiation where the immigrant would care about the origin of the product and shape preferences around the variety of the good. The impact of immigrants on reduction in fixed costs can also be in operation here whereby immigrants provide specialised information about markets in their home country. But for differentiated products, the fixed costs entails much more complex transactions that are harder to overcome than those for more homogeneous products, so immigrants have less of an impact on costs for these products compared with more homogeneous ones. However, since the impact on immigration is not statistically significant on differentiated exports, I take this to be suggestive evidence that the preferences channel could be the more prominent channel.

For Mercosur and newly arriving immigrants, the positive significant impact on exports at low levels of differentiation, which is also at the product extensive margin, is suggestive that a channel of impact at play is immigrants' ability to reduce fixed costs for such products. The model of Tai (2009) predicts that immigration reduces fixed costs more for goods who are more homogeneous and at less levels of differentiation. This is because those products are presumably easier and more straightforward to trade, relative to differentiated products. However, this impact through fixed costs on homogeneous products translates into lower effects on exports. because those types of products are highly substitutable, and hence trading firms will only be able to capture a small portion of the market. And this is indeed what I find whereby a doubling of the average of Mercosur migrants per population across all municipalities increases homogeneous exports by a modest 0.018%.

In this section, I have estimated the causal impact of immigration on trade. Next, I estimate the impact of immigration at the country-municipality level, to estimate results on the heterogeneous impacts of the role of immigrants on trade based on characteristics of the country of origin.

Table II.8: Total Exports and Imports

	(1) log(Total Exports)	(2) log(Total Exports)	(3) log(Total Imports)	(4) log(Total Imports)
Migrants	-0.0146	0.0854	0.00691	0.337
	(0.0127)	(0.275)	(0.00950)	(0.231)
Observations	44543	44543	44543	44543
Estimation	OLS	IV	OLS	IV
Mean Migrants	0.000633	0.000633	0.000633	0.000633
Doubling of Mean	-0.000922	0.00541	0.000437	0.0214
Municipality Controls	Yes	Yes	Yes	Yes
Year F.E	Yes	Yes	Yes	Yes
Municipality F.E	Yes	Yes	Yes	Yes

Notes: The table shows the OLS and 2SLS estimates for the impact of immigration on exports and imports. The dependent variable is of the form $log(x+1)$ and the explanatory variable of immigration is the ratio of immigrants to the population per municipality (in %). Robust standard errors in parentheses clustered at the mesoregion level. Regressions include mesoregion fixed effects and municipality controls. The predicted migration variable is divided by population in each municipality. Data sources: IBGE data for control variables, COMEX for trade data and SINCRE for migration data. Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1.

Table II.9: Differentiated and Homogeneous Exports and Imports based on Rauch Classification

	(1) log(Differentiated Exports)	(2) log(Differentiated Exports)	(3) log(Homogeneous Exports)	(4) log(Homogeneous Exports)
Migrants	-0.00969	-0.133	-0.0149	0.212
	(0.0107)	(0.258)	(0.0132)	(0.228)
Doubling of Mean	-0.000613	-0.00839	-0.000942	0.0134
	(1) log(Differentiated Imports)	(2) log(Differentiated Imports)	(3) log(Homogeneous Imports)	(4) log(Homogeneous Imports)
Migrants	0.00307	0.431*	-0.00856	0.0382
	(0.00536)	(0.228)	(0.0140)	(0.208)
Doubling of Mean	0.000194	0.0273	-0.000541	0.00242
Observations	44543	44543	44543	44543
Estimation	OLS	IV	OLS	IV
Mean Migrants	0.000633	0.000633	0.000633	0.000633
Municipality Controls	Yes	Yes	Yes	Yes
Year F.E	Yes	Yes	Yes	Yes
Municipality F.E	Yes	Yes	Yes	Yes

Notes: The table shows the OLS and 2SLS estimates for the impact of immigration on differentiated and homogeneous exports. The dependent variable is of the form $log(x+1)$ and the explanatory variable of immigration is the ratio of immigrants to the population per municipality (in %). Robust standard errors in parentheses clustered at the mesoregion level. Regressions include mesoregion fixed effects and municipality controls. The predicted migration variable is divided by population in each municipality. Data sources: IBGE data for control variables, COMEX for trade data and SINCRE for migration data. Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1.

Table II.10: **Extensive Margin:** Differentiated and Homogeneous Exports and Imports based on Rauch Classification

	(1) log(Total Exports)	(2) log(Differentiated Exports)	(3) log(Homogeneous Exports)	(4) log(Total Imports)	(5) log(Differentiated Imports)	(6) log(Homogeneous Imports)
Migrants	0.0155	0.147	0.322	0.0962	-0.0664	-0.180
	(0.284)	(0.252)	(0.280)	(0.227)	(0.218)	(0.203)
Observations	44543	44543	44543	44543	44543	44543
Estimation	IV	IV	IV	IV	IV	IV
Doubling of Mean	0.000978	0.00931	0.0204	0.00608	-0.00420	-0.0114
Municipality Controls	Yes	Yes	Yes	Yes	Yes	Yes
Year F.E	Yes	Yes	Yes	Yes	Yes	Yes
Municipality F.E	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The table shows the OLS and 2SLS estimates for the impact of immigration on differentiated and homogeneous imports. The dependent variable is of the form $log(x+1)$ and the explanatory variable of immigration is the ratio of immigrants to the population per municipality (in %). Robust standard errors in parentheses clustered at the mesoregion level. Regressions include mesoregion fixed effects and municipality controls. The predicted migration variable is divided by population in each municipality. Data sources: IBGE data for control variables, COMEX for trade data and SINCRE for migration data. Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1.

Table II.11: **Intensive Margin:** Differentiated and Homogeneous Exports and Imports based on Rauch Classification

	(1) log(Total Exports)	(2) log(Differentiated Exports)	(3) log(Homogeneous Exports)	(4) log(Total Imports)	(5) log(Differentiated Imports)	(6) log(Homogeneous Imports)
Migrants	0.000809	-0.0580	-0.130	0.343*	0.480**	0.197
	(0.220)	(0.199)	(0.198)	(0.201)	(0.227)	(0.195)
Observations	44543	44543	44543	44543	44543	44543
Estimation	IV	IV	IV	IV	IV	IV
Doubling of Mean	0.0000512	-0.00367	-0.00822	0.0217	0.0304	0.0124
Municipality Controls	Yes	Yes	Yes	Yes	Yes	Yes
Year F.E	Yes	Yes	Yes	Yes	Yes	Yes
Municipality F.E	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The table shows the OLS and 2SLS estimates for the impact of immigration on differentiated and homogeneous imports. The dependent variable is of the form $log(x+1)$ and the explanatory variable of immigration is the ratio of immigrants to the population per municipality (in %). Robust standard errors in parentheses clustered at the mesoregion level. Regressions include mesoregion fixed effects and municipality controls. The predicted migration variable is divided by population in each municipality. Data sources: IBGE data for control variables, COMEX for trade data and SINCRE for migration data. Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1.

Table II.12: Based on σ Elasticity of Substitution

	(1) log(Exports Low Diff.)	(2) log(Exports Low Diff.)	(3) log(Exports Mid Diff.)	(4) log(Exports Mid Diff.)	(5) log(Exports High Diff.)	(6) log(Exports High Diff.)
Migrants	-0.00925	0.110	-0.0184	-0.0299	0.00963	0.238
	(0.0122)	(0.285)	(0.0164)	(0.278)	(0.0107)	(0.250)
Observations	44543	44543	44543	44543	44543	44543
Estimation	OLS	IV	OLS	IV	OLS	IV
Mean Migrants	0.000633	0.000633	0.000633	0.000633	0.000633	0.000633
Doubling of Mean	-0.000586	0.00699	-0.00117	-0.00189	0.000610	0.0151
	(1) log(Import Low Diff.)	(2) log(Import Low Diff.)	(3) log(Import Mid Diff.)	(4) log(Import Mid Diff.)	(5) log(Import High Diff.)	(6) log(Import High Diff.)
Migrants	0.0126	0.328	0.00416	0.330	-0.0143	0.211
	(0.0120)	(0.233)	(0.00834)	(0.216)	(0.0137)	(0.226)
Doubling of Mean	0.000796	0.0208	0.000263	0.0209	-0.000905	0.0134
Municipality Controls	Yes	Yes	Yes	Yes	Yes	Yes
Year F.E	Yes	Yes	Yes	Yes	Yes	Yes
Municipality F.E	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The table shows 2SLS results for subcategories of exports and imports based on the elasticity of substitution σ. Higher elasticity of substitution means that the product is less differentiated and hence more homogeneous. The dependent variable is of the form $log(x+1)$ and the explanatory variable of immigration is the ratio of immigrants to the population per municipality (in %). Robust standard errors in parentheses clustered at the mesoregion level. Regressions include mesoregion fixed effects and municipality controls. The predicted migration variable is divided by population in each municipality. Data sources: IBGE data for control variables, COMEX for trade data and SINCRE for migration data. Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1.

Table II.13: **Extensive Margin:** Based on σ Elasticity of Substitution

	(1) Exports Low Diff.	(2) Exports Mid Diff.	(3) Exports High Diff.	(4) Import Low Diff.	(5) Import Mid Diff.	(6) Import High Diff.
Migrants	0.465	0.0251	0.335	-0.0357	0.250	0.0892
	(0.304)	(0.254)	(0.238)	(0.233)	(0.245)	(0.230)
Observations	44543	44543	44543	44543	44543	44543
Estimation	IV	IV	IV	IV	IV	IV
Doubling of Mean	0.0294	0.00159	0.0212	-0.00226	0.0158	0.00564
Municipality Controls	Yes	Yes	Yes	Yes	Yes	Yes
Year F.E	Yes	Yes	Yes	Yes	Yes	Yes
Municipality F.E	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The table shows 2SLS results for subcategories of exports and imports based on the elasticity of substitution σ. Higher elasticity of substitution means that the product is less differentiated and hence more homogeneous. The dependent variable is of the form $log(x+1)$ and the explanatory variable of immigration is the ratio of immigrants to the population per municipality (in %). Robust standard errors in parentheses clustered at the mesoregion level. Regressions include mesoregion fixed effects and municipality controls in 2010. The predicted migration variable is divided by population in each municipality. Data sources: IBGE data for control variables, COMEX for trade data and SINCRE for migration data. Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1.

Table II.14: **Intensive Margin:** Based on σ Elasticity of Substitution

	(1) Exports Low Diff.	(2) Exports Mid Diff.	(3) Exports High Diff.	(4) Import Low Diff.	(5) Import Mid Diff.	(6) Import High Diff.
Migrants	-0.0179	0.102	0.285	0.396*	0.169	0.239
	(0.245)	(0.233)	(0.189)	(0.221)	(0.193)	(0.182)
Observations	44543	44543	44543	44543	44543	44543
Estimation	IV	IV	IV	IV	IV	IV
Doubling of Mean	-0.00113	0.00643	0.0181	0.0251	0.0107	0.0151
Municipality Controls	Yes	Yes	Yes	Yes	Yes	Yes
Year F.E	Yes	Yes	Yes	Yes	Yes	Yes
Municipality F.E	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The table shows 2SLS results for subcategories of exports and imports based on the elasticity of substitution σ. Higher elasticity of substitution means that the product is less differentiated and hence more homogeneous. The dependent variable is of the form $log(x+1)$ and the explanatory variable of immigration is the ratio of immigrants to the population per municipality (in %). Robust standard errors in parentheses clustered at the mesoregion level. Regressions include mesoregion fixed effects and municipality controls in 2010. The predicted migration variable is divided by population in each municipality. Data sources: IBGE data for control variables, COMEX for trade data and SINCRE for migration data. Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1.

Heterogeneous Impacts by Country of Origin Characteristics In what follows, I estimate the impact of immigration on trade at the country and municipality levels, using the same instrumental variable approach. It is likely that the overall effect of immigration on trade indicated in previous sections masks large heterogeneity depending on their country of origin characteristics.[41] In this section, I explore particularly this angle by disaggregating the analysis based on key country of origin characteristics. While immigrants of all origins can alter preferences of the host community, or reduce fixed costs through their networks and connects, their ability to do so effectively and successfully will depend on the institutional set-up, background, level of development among others, in their country of origin. The three country of origin characteristics are likely to play a role in facilitating how immigrants impact trade: GDP level, ethnic fractionalisation, and interpersonal trust, however, it is worthy to note that it is by no means an exhaustive list of country of origin characteristics that are relevant and important for trade.[42]

I choose these three country of origin characteristics for the following reasons. First, migrants coming from countries with different levels of development might have differ-

[41] As a corollary of this, an interesting point is whether the migrant, regardless of their country of origin, impact trade differently depending on individual characteristics such as education, gender, or race. The data I have on the migrants' education or occupation level is noisy and inaccurate so I cannot explore fully this angle using this dataset.

[42] Importantly, the type of immigrant, their intention to stay in the host country and whether they are voluntary or economic immigrants, can differ based on the country of origin characteristics and so the results are confounding these two aspects together.

ent impacts on trade. This could be due to the level of development of the institutions of the country of origin, ease of doing business and established trade relations which are highly correlated with level of development measured by GDP levels. Immigrants from countries with less developed institutions might actually play a more prominent role in their impact on trade because of their ability to overcome the initial large fixed trade costs (Peri, 2010; Briant, Combes and Lafourcade, 2009). I proxy for the level of institutional development using the level of GDP averaged from 1985-2006 (so prior to the period studied). I obtain this measure from the dataset provided by Ashraf et al. (2020).

Second, I study whether more diverse communities affect trade differently than those that tend to be more homogeneous. The underlying hypothesis behind this measure is that migrants coming from more ethnically diverse communities would possess an aptitude to building connections across cultures and hence would be more likely to impact trade positively as opposed to migrants coming from homogenous communities. To proxy this measure, I use the index of ethnic fractionalisation. This index reflects the probability that if two individuals were selected randomly from a country, they will belong to different ethnic groups.[43]

Third, I explore whether the prevalence of inter-temporal trust in the country of origin for the migrants ameliorates the impacts that immigrants have on trade. As Arrow noted, "virtually every economic transaction has an element of trust" (Arrow, 1972). Trust is very important in reducing transaction costs, enforcing contracts and facilitating payment. I proxy for this measure using the variable of interpersonal trust in Ashraf et al. (2020) which is constructed using the World Values Survey (2006, 2009). This measure reflects the proportion of respondents that opted for the answer "Most people can be trusted" when answering the following survey question: "Generally speaking, would you say that most people can be trusted or that you need to be very careful in dealing with people?".

In order to be able to estimate the 2SLS equation for immigrants belonging to different countries of origin based on a certain characteristic, I estimate the following equation at the level of the country of immigrant, municipality and year and adjust the shift share IV accordingly to match the functional form of the estimation in equation

[43]Formally, the index is calculated as follows: ethnic fractionalisation $= 1 - \sum_{j=1}^{n} p_i^2$ where p_i is the proportion of ethnicity i to the population in the country and n is the total ethnic groups (Alesina el al. 2003).

(5.21) as described below.[44]

$$X_{i,j\in z,t} = \upsilon_{i,t} + \gamma_{i,j} + \delta mig_{i,j\in z,t} + \nu\chi_{i,t} + \epsilon_{i,j,t}. \tag{5.21}$$

In equation 5.15, j represents each country, and z represents a certain country group characteristic. $X_{i,j\in z,t}$ is a measure of exports or imports between municipality i and the countries j belonging to a certain range of the value of a country characteristic z, $\upsilon_{i,t}$ is year fixed effects, $\gamma_{i,j}$ is the municipality fixed effect, $mig_{i,j\in z,t}$ measures the number of migrants as a share of the population belonging to the country j that has a certain value of country characteristic that lies within the range z, $\chi_{i,t}$ is as before vector of municipality controls and $\epsilon_{i,j,t}$ is the error term. So I estimate separate equations, each one pools together the countries that are within a certain percentile range of a characteristic about their country of origin[45]. The instrumental variable is the as described in equation (5.20).

It is worth to note, however, that in the estimation for the different sample of the country groups, the IV 2SLS first stage Kleibergen Paap (K-P) F statistic falls significantly below 10. This is indicating that at this level of estimation, the instrument is weak and hence this should be kept in mind when interpreting the results. Whenever this is the case, I mention it in the discussion of the results.

I present the estimates of the impacts on trade due to a 1% increase from the mean of the immigrant ratio to the population, disaggregating countries by the percentiles of each characteristic, in figure II.6.[46] Interestingly, I find that countries with

[44]Peri and Riquena-Silvente (2010) and Ehrhart (2014) use a similar methodology in an attempt to get to a causal estimate of migration on trade and to understand better the role of the country of origin. Another method to estimate this equation is by pooling countries that belong to certain level of the country of origin characteristic discussed and estimate the equation at the municipal level. I test this approach and I find the results to be qualitatively consistent with the findings reported in this section. However I prefer the approach discussed here because it allows me to link the immigrants to trade to and from their country of origin. As an addition, it allows a nice link to the empirical estimation in section 4.2.

[45]For example, the estimation is done at the migrant's country of origin, municipality and year level. For every characteristic of the country of origin studied, like GDP levels, I estimate equation 5.15 on all countries that are within a certain percentile range of that characteristic (for example, those that have a GDP level up to 25$^{\text{th}}$ of the sample.). I also attempt to estimate equation (5.21) but also adding country-year fixed effects. I find that I lose a lot of power in the first stage results of the IV 2SLS because adding this extra fixed effects makes me lose a lot of variation.

[46]The tables for these results in the appendix also report the impact on trade due to a one percentage point change in immigration. Note that some of these reported estimates are very large. It worth to keep in mind that they reflect the impact on trade due to a one percentage point increase in the ratio of migrants to population. In certain cases, depending on the migrant community, a one percentage point increase might mean a 2,906% increase from the average. The average of the migrant to population differs across the estimations because it depends on the country and municipality pair trading certain classification of goods.

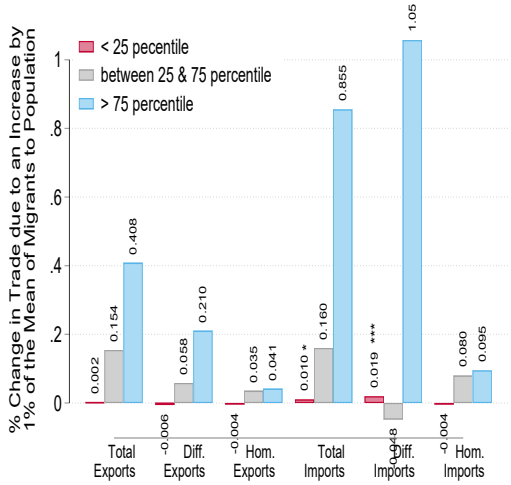
a GDP level below the 25th percentile of the GDP levels of all countries of origin of immigrants, causes higher levels of differentiated imports. Increasing the mean of immigrants from those countries by 1%, increases differentiated imports by 0.02%. By contrast, immigrants arriving from countries at the upper tail of the GDP distribution have an insignificant correlation with trade. These findings could be due to the fact that immigrants in less developed economies are able to contribute more to overcome the institutional bottlenecks and fixed costs associated with differentiated products leading to an increase in trade, that would have been much harder to happen without the presence of the specialised knowledge of an immigrant community.

I find that migrants coming from more homogeneous communities, lower ethnic fractionalisation, have a significant and positive impact on total imports as well as differentiated and homogeneous imports. The point estimates reported, although statistically significant, are potentially not economically significant (for example, an increase of 0.003% of differentiated imports in response to increasing the mean by 1%). Whereas migrants coming from countries with high levels of fractionalisation (above the 75th), have a negative impact on differentiated exports and imports. For example, differentiated exports would fall by 0.024% if the mean of immigrants from those countries were to increase by 1%. Furthermore, I find that immigrants coming from levels of interpersonal trust below the 25th percentile have no impact on trade, while those that come from countries with higher levels of interpersonal trust (above 75th percentile) have a positive impact on homogeneous exports as well as total, differentiated and homogeneous imports. The channels of impact that could be at play are potentially changes in preferences or reduction in fixed costs as evidenced by the resulting increase in homogeneous and differentiated imports.[47].

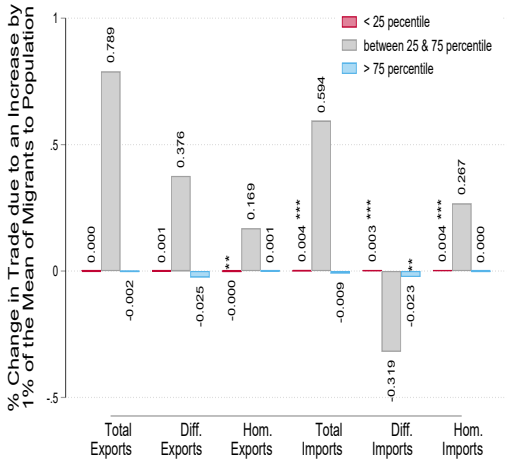
[47]I also test whether the years under democracy or autocracy as well as levels of political opinion dispersion in the country of origin affects immigrants' ability to influence and affect trade. I find that those variables seem to play a minimal to no role in the heterogeneity of impact.

Figure II.6: Heterogeneous Impact on Trade by Country of Origin Characteristics

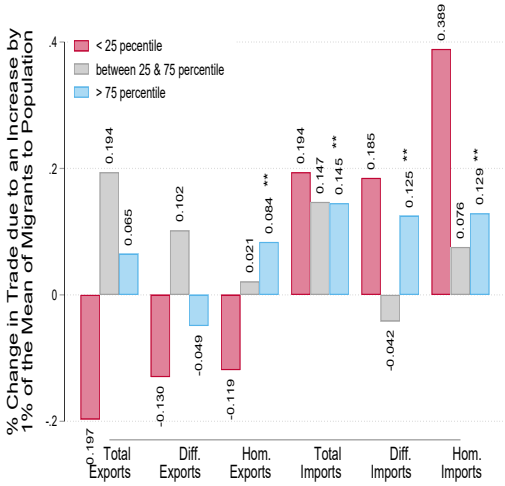
(a) GDP Levels



(b) Ethnic Fractionalisation



(c) Prevalence of Interpersonal Trust



Notes: These figures show the 2SLS second stage estimations of the impact of immigrants on trade. The immigrants are disaggregated by the level of GDP, Ethnic Fractionalisation and interpersonal trust, into those up to the 25th percentile, between the 50th and 75th percentiles, and those above the 75th percentile. The estimations are shown for total imports and exports, and also disaggregated by differential and homogeneous goods. The point estimates are also written next to each bar and I indicate using asterisks those that are statistically significant from zero as such for p-values: *** p<0.01, ** p<0.05, * p<0.1. Data sources: IBGE data for control variables, COMEX for trade data and SINCRE for migration data.

6 Limitations of the Analysis

In this section, I discuss the limitations that this analysis faces and concerns pertaining to the exclusion restriction of the instrumental variable strategy.

Trading companies in Brazil seem to be concentrated in the South and South-East regions of Brazil (as can be seen from figure II.3(b)). Those regions are also the ones that are more developed than other regions and ones that I would expect to attract more economic immigrants.[48] This raises some concerns over the potential importance of internal trade between municipalities within Brazil. With this context, one might be concerned that migrants living, say in the North, through their own demand or influence of preferences of the natives, would also impact trade through established firms in the South. If these firms are based in municipalities with no immigrants, then my estimates will reflect a lower bound of the estimated relationship, i.e the impact would be more diffused. Hence, because I am not able to identify the final destination of consumption of the imported goods, the coefficients estimated on the impacts on imports is subject to attenuation bias. With constraints on data availability, I cannot quantify internal trade or evaluate how it would influence the results. But this is an important channel that can be considered in future work. Another issue that that the literature has considered is the role of the extensive and intensive margins of trade at the firm level. In particular, the literature considered the extensive margin of trade in terms of the entry of new trading firms, while the intensive margin is in terms of an increase in transactions for the same firm. This level of analysis will allow further evidence on the channels at play. At this stage, and due to data availability, I cannot provide such an analysis and can only study the intensive and extensive measures at the product and country level. This analysis allows me to understand in more depth if the presence of immigrants was also associated with information about new products that were not being traded prior to their arrival.

There are also limitations with respect to the instrumental variable strategy. The first is regarding the language assigned to each country. In this analysis, each country is assigned the language spoken by the majority of the citizens as per the CIA The World Factbook and World Atlas Language Structures. However, individuals from a certain country may speak different languages. Also, one should expect that migrants are self selected and hence more likely to be speaking more than one language or a Lingua Franca. Another limitation, is that the estimation of the IV 2SLS assumes

[48] Those are immigrants who migrate for the main purpose of seeking better economic and job market opportunities in the host country.

that the locations are independent and that there are no spillovers. This assumption may not hold in this context because spatial spillovers can exist through changes in commuting patterns due to local labour demand shocks (Monte, Redding, and Rossi-Hansberg, 2018). To limit potential spillovers, through the channels presented such as information, preferences and network, between firms, nationals and migrants, I take the unit of analysis to be the municipality and I perform different robustness checks at two higher levels of fixed effects (Federation Unit and mesoregion level). While I prefer the comparison at the municipality level, given that it is the most disaggregated spatial distribution of trade flows and immigration, and potentially the strictest level of fixed effects in terms of finding any impact, I present, in tables II.38 to II.45, the results using mesoregion and federation unit fixed effects. The results are broadly similar with the exception that in the specification where fixed effects are the federation unit level interacted with years, the presence of migrants has a positive and significant impact on total exports. This seems to be driven by an increase in the differentiated exports.

An important argument in the IV shift share literature is presented in Jaeger et al. (2018). The authors point out that adjustment dynamics are important. The IV shift share instrument is meant to be exogenous to local demand or supply shocks that affect both trade and migrants' location preferences. However, markets will take time to adjust to past shocks and this will be captured in the contemporaneous error terms. Therefore, the IV estimate will conflate the short term response of new migrants and the long term response, that for example could capture capital adjustment. In order to see if I am able to account for short run as well as long run dynamics, I calculate the pair-wise correlations across migrant cohorts by language for all years under study and present the results in tables II.19 and II.20. I notice that there seems to be strong correlation across cohorts indicating that the language composition of cohorts changed very little across the years. Therefore the instruments used to identify the exogenous variation in the two migrant ratios will be highly correlated and hence, my first stage estimation will suffer from a joint weak instrument problem in finite samples. The multiple instrumentation approach developed by Jaeger et al. (2018) is very important but demanding on the data mainly because it requires that the compositional change in migrants cohorts to vary substantially across time. This is not the case in this analysis within the time frame I consider (2010-2017) and hence I will not be able to separately identify short run and long run dynamics.

Finally, I use in this analysis all the municipalities in Brazil where data is available. This means that there are many zeroes in municipalities that do not trade. Therefore,

I reproduce the results for the sample of municipalities that traded in the period 2010-2017. I also perform robustness checks where I exclude municipalities that have a volume of trade above the 90^{th} percentile, or municipalities that are at the border. I present those results in the appendix and I find that consistent and estimates that are similar qualitatively to main specifications provided in the text.

7 Conclusion and Policy Implications

In this chapter, I examined the impact of immigration on trade from 2010 to 2017 in the Brazilian context, across products at different levels of differentiation. First, I discuss the channels through which immigrants can impact trade at different degrees of product differentiation and provide a simple empirical exercise based on Tai (2009)'s model to link the channels to the data. Second, I estimate the relationship at a country, municipality and year level motivated by the augmented gravity equation from the model using a rich set of fixed effects. Third, since locations of immigrants in Brazil are likely to be endogenous to local labour demand factors, I employ a variation on the shift share instrument to predict exogenous variation in migrant locations based on earlier settlements of migrant groups in Brazil that share a similar language to new migrants. And fourth, I employ the IV shift share at the country and municipality levels to estimate the heterogeneous impacts on trade depending on the characteristics of the country of origin.

Based on the estimation motivated by the augmented gravity equation (Tai, 2009), I find that the presence of immigrants is correlated positively with imports at low and high levels of product differentiation. This effect is also found at the product extensive margin. The results on the 2SLS IV municipality-year level, show that migrants increase differentiated imports. I link each immigration stock to trade with their country of origin and estimate a causal impact of immigration on trade using the same instrumental variable strategy. My results are consistent with what I find with the estimation at the municipality level, where I find an increase in total and homogeneous imports and imports at low and medium levels of differentiation, at the product extensive level. The results on differentiated imports suggest that migrants are impacting both preferences (i.e. migrants affecting the preferences of other country residents and/or demanding country of origin specific products) and fixed trade costs, with the former being more prominent than the latter because I do not observe

a similar positive impact on differentiated exports. Whereas the impacts on homogeneous goods could be indicative of immigrants reducing trade costs that are easier to overcome for such products. The presence of an impact on such goods on the extensive margin (immigrants are bringing *new* products) is also supportive of the hypothesis that immigrants are able to overcome trade costs, especially those costs that are can be more straightforward for less differentiated products. My analysis is complemented with evidence that the impact on trade and by conjunction the channels through which immigrants affect trade vary depending on the characteristics of the country of origin. I study three characteristics of the country of origin, such as the level of development, ethnic fractionalisation, and prevalence of interpersonal trust. I find that immigrants coming from countries with low levels of ethnic fractionalisation increase imports, both homogeneous and differentiated, while those with higher levels of ethnic fractionalisation lead to less imports, especially differentiated imports. I find that the latter result holds for immigrants coming from countries that are characterised with a higher level of interpersonal trust. Interestingly, I also report a significant increase in differentiated imports due to an increase in the presence of immigrants from countries at the lowest end of the countries' distribution of GDP levels (less than 25^{th} percentile of the sample) and an increase in differentiated exports for immigrants coming from countries of origin with a GDP level lying within the 25^{th} and 75^{th} percentile of the sample distribution. This impact could be explained by the role that immigrants play in overcoming weak institutions or trade barriers in their country of origin, both of which can be correlated with lower levels of development (Ehrhart et al. 2014).

The question that this paper addresses has both policy relevance and policy implications. Migration is a salient political debate and the trade-offs that migrants bring to bear on the host community are key in impacting how and if migrants are welcomed and how they are assimilated into the economy. This is especially important considering the reluctance of many countries to welcome refugees and forced migrants. Therefore understanding how migrants impact trade in the host economy is important, especially in the context of South-South migration. Further, migration episodes are to be expected to continue to happen, especially refugee migration due to climate change crises or political violence. Further, my analysis focuses on a period where immigration into Brazil increased significantly, and where immigration began from countries that did not have any or meagre historic presence in Brazil. Therefore, the reported pro-trade creation effect on homogeneous and less differentiated goods due to the presence of immigration could be related to this context of relatively new and accelerated

immigration, where it could be easier for those immigrants to overcome the less complicated fixed costs associated with such projects. It will be interesting to study how how the impacts on trade will differ as immigrants become more established in the next decades.

In terms of policy implications that arise from this study, I find that immigration presents the host country with trade-offs. On one hand, migrants contribute to increases in exports, but may also pose a humanitarian fiscal burden on the government especially in the case of recent waves of Haitians and Venezuelans (considered to be humanitarian refugees) or competition in the labour market, in the case of migrants from other countries with higher levels of development. The impact of migration on the last two channels goes beyond the scope of this paper, however, it is important to place the results from the impact on trade within this general framework to inform potential distribution policies for the receiving country. Further, the positive impact on differentiated exports (in the case of an increase in immigrants at intermediate GDP levels) may also increase the acquisition of valuable capabilities (many anchored in market-based knowledge) and contribute to the growth of the host country.

The analysis presented in this paper could be extended. The paper indicated that the trade creation effect of immigrants differs based on the country of origin based on three country of origin characteristics studied in this empirical analysis. However, this is far from being an exhaustive list and future work can include other characteristics that vary by the country of origin, and ones that are at an individual level and differ based on the immigrant's level of education or type of profession. Studying the impacts of migrants at the firm level might offer another important avenue for research. In an attempt to ground the analysis within a conceptual framework, this paper presented a simple empirical exercise to guide the empirical work in understanding how can the impact of immigration on trade be realised at different levels of product differentiation. Nonetheless, while the latter exercise is useful in providing plausible interpretation for the results, the indicated channels remain descriptive and therefore incorporating the pro-trade creation effects of immigrants, in a structural model, can help disentangle the channels more accurately. This chapter underscores the point that there is much yet to learn about the impacts of immigration in an economy of the global South.

Acknowledgements I would like to thank Professor Carlos Vargas Silva, the coordinators of the migration initiative at the Inter-American Development Bank (IDB), the workshop Understanding the Impacts of Migration in Latin America and the Caribbean at the IDB, the Oxford Latin American Centre seminar series, and the ODID Quantitative Workshop. I would also like to thank Leonardo Cavalcanti, Tadeu Oliveira, Marília Fernandes and all the staff of the Observatório das Migrações Internacionais (OBMigra) at the Universidade de Brasília for providing access to the data and answering multiple questions about it. This work is based on the SINCRE edition of the Immigration dataset when it was first released.

Appendices of Chapter 2

II.A Supplementary Material: Validity Tests of the Shift Share IV

The IV shift share instrument aggregates different underlying instruments, which in my case are different language groups residing in Brazil up to the year 2000. Recent literature surrounding IV shift share strategies encouraged researchers to understand where the sources of variation in the IV are coming from.

Following the approach outlined by Goldsmith-Pinkham, Sorkin, and Swift (2019). Goldsmith-Pinkham et al. (2019) show that in finite samples, the IV 2SLS estimator is numerically equivalent to a generalised method of moments (GMM) estimator where the latter is a weighted sum of just-identified estimates based on a finite sample decomposition due to Rotemberg (1983). Through recasting the IV estimator of the Bartik instrument as a GMM estimator, the IV estimator can be decomposed into a set of estimators each representing the local shares and a set of Rotemberg weights associated with each instrument (I refer readers to Goldsmith-Pinkham, Sorkin, and Swift (2019) for details on this approach). Each of the just identified estimates with their associated Rotemberg Weights will provide the contribution of each share to the overall Bartik IV 2SLS estimate.

The assumption behind the IV that I construct in this analysis is that municipalities have different exposure to common shocks of migrants' flows based on earlier immigration settlements. Therefore the exogeneity assumption is that the historic language shares of migrants at time $t = 0$ are exogeneous to trade outcomes in the period of analysis 2010-2017. However it is important to understand two key aspects: 1- which language groups are the most important in explaining the new immigrant locations and 2- whether those language groups were able to predict baseline outcomes that can be associated with changes in trade in 2010-2017. I present the results in table III.15. This decomposition is for all municipalities, all languages and for years 2011,2013, 2015, and 2017. Panel A summarises the weights α_f decomposed by positive and negative Rotemberg weights. One can see that the positive weights constitute the highest share of overall weight. Panel B explores the correlation between these shares, the national migrant flows, the just identified estimates β_f, the first stage F-test and the variance of the shares of this instrument z_f. The Rotemberg weights are positively correlated with the national migrant flows and the variance of the local language shares. Higher dispersion in the latter implies a higher variation in the cross sectional exposure to migrants who speak different languages, and hence creating higher variation in the

instrument. Also, more variation in the national flows of migrants speaking a certain language, increases the variation in the instrument. As shown in Panel C is that the variation in the IV instrument is mainly being driven by the presence of Spanish speakers at $t = 0$ where it accounts for 81% (0.819/1.001) of the positive weight of the estimator.

Due to the high weight on the Spanish share, one should examine it more carefully. I also study whether the original distribution of migrants speaking the main language groups identified in the aforementioned exercise are plausibly exogenous to shocks affecting trade.I regress the Bartik IV and the language shares on observable characteristics for the year 2000 ($t = 0$) by estimating the following equation and report the results in table II.16:

$$Z_i = \alpha + \gamma_1 X_{i,0} + \theta_r + \epsilon_m. \tag{II.A.22}$$

where Z_i is the Bartik IV or the individual language shares of $\in Spanish, Japanese, Hindu, Italian, Haitian$. Studying the correlation for the initial time period is a tighter restriction and reflects the instrument's cross sectional variation. If the shares at time $t = 0$ are correlated with potential confounding factors, then this might be indicative that there are other omitted variables biasing the estimation. As seen in table II.16 there is weak to no correlation between the shares and variables potentially correlated with trade, Also the R-squared ranges between 0.1 to 0.26 are consistently small suggesting that languages contributing to the most variation are not systemically located in certain municipalities in Brazil. However, there is a significant and positive correlation between early Spanish settlements and the difference in value added in agriculture and poverty rate between 2010 and 2000. This is indicative that earlier Spanish communities located in areas that became poorer and more concentrated in agriculture. This finding poses a threat to the exclusion assumption of the instrument if those municipalities where also ones that developed better trade relations or affected factors influencing trade. The lack of significance on the correlation between Spanish early immigrants and other variable such as per capita GDP, value added in industry and value added in services is reassuring that these municipalities did not necessarily evolve in ways that favoured trade relations in the future[49]

One can argue that immigrants located strategically into areas that had prospering trade or had favourable factors to future economic growth. Therefore as another test, I

[49]I also perform this test using a specification that regresses each of the covariates on the language shares and I find the results to hold.

regress differences in the initial period covariates (from 2010 to 2000, as recommended by Goldsmith et al. 2018) on migration concentration across municipalities in the subsequent decade. I report the results in table II.17 and I find that, reassuringly, there is no correlation between future immigrant location decisions and beginning of period covariates that could be potentially important for trade growth.

And finally, it remains important to test if the historic language shares are correlated with pre-trends of trade leading up the period of analysis. I estimate the following equation:

$$X_{i,t} = \sum_{T=2001(4)2009} [\sigma_T(L_i) \times (t = T)] + \theta_{i,t} + \epsilon_{i,t}, \tag{II.A.23}$$

where $\theta_{i,t}$ is municipality-year fixed effects. I report the results for total imports and exports in table II.18 and I find that the historic language settlements are uncorrelated with pre-trade trends[50].

Table II.15: Evaluation of Underlying Shares

Panel A: Negative and positive weights

	Sum	Mean	Share
Negative	-0.001	-0.000	0.001
Positive	1.001	0.013	0.999

Panel B: Correlations of Language Groups

	α_f	g_f	β_f	F_f	Var(z_f)
α_f	1				
g_f	0.963	1			
β_f	0.033	0.054	1		
F_f	0.040	0.034	0.026	1	
Var(z_f)	0.195	0.182	0.111	-0.027	1

Panel C: Top 5 Rotemberg weight industries

	$\hat{\alpha}_f$	g_f	$\hat{\beta}_f$
Spanish	0.819	73669.000	0.011
Italian	0.028	3997.000	-0.005
Japanese	0.027	2443.000	-0.002
Hindi	0.027	2361.000	-0.002
Haitian	0.019	21278.000	0.046

Notes: This table provides an analysis on the underlying shares forming the shift-share instrumental variable approach presented in this paper (Goldsmith-Pinkham, Sorkin and Swift, 2019). α_f represents the Rotemberg Weights calculated for every language which represent teh contribution of each share to the overall Bartik IV 2SLS estimate; g_f represents the national flows of migrants who speak language f; β_f represents the contribution of each instrument to the overall Bartik instrument and z_f represents the IV for every language f.

Table II.16: Correlations with the Highest Rotemberg Language Shares and Potential Covariates

	(1) IV	(2) Spanish	(3) Japanese	(4) Hindu	(5) Italian	(6) Haitian
Diff(per Capita GDP)	-0.000000787	-0.00000184	-0.000000142	-1.92e-08	-0.000000433	-0.000000127
	(0.000000560)	(0.00000135)	(0.000000254)	(6.30e-08)	(0.000000343)	(0.000000117)
Diff (VA agriculture)	0.0223***	0.0535***	0.00329	0.00159	0.00534	0.00121
	(0.00275)	(0.00933)	(0.00361)	(0.00118)	(0.00434)	(0.000748)
Diff (VA services)	0.00128	0.00179	0.00331	0.00101	-0.00152	0.000425
	(0.00296)	(0.0101)	(0.00422)	(0.00145)	(0.00441)	(0.000388)
Diff (VA industry)	0.00201	0.00899	0.00554	-0.000466	0.00231	-0.000627
	(0.00274)	(0.00939)	(0.00452)	(0.000693)	(0.00548)	(0.000575)
Diff (Taxes)	-0.00970***	-0.0333***	-0.00325	-0.000177	-0.00728	-0.000433
	(0.00343)	(0.0116)	(0.00476)	(0.000595)	(0.00469)	(0.000999)
Diff(GINI index)	-0.0256	-0.0672	-0.0614	0.00387	-0.0697	0.00190
	(0.0294)	(0.0996)	(0.0375)	(0.00608)	(0.0434)	(0.00400)
Diff(Average Salary)	0.00000270	0.00000243	0.000000409	0.000000426	0.00000342*	-0.000000590
	(0.00000283)	(0.0000107)	(0.00000211)	(0.000000299)	(0.00000205)	(0.000000442)
Diff(Poverty Rate)	0.0233*	0.101**	-0.0520**	0.00322	-0.00552	-0.00282
	(0.0128)	(0.0488)	(0.0254)	(0.00313)	(0.0218)	(0.00308)
Observations	2695	2695	2695	2695	2695	2695
Adjusted R sq	0.2661	0.1819	0.0617	0.1031	0.0615	0.1069
Microregion F.E	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This figure shows correlation between the shares across the 5 languages that had the highest Rotemberg weights with the initial period covariates. The covariates include log differences in percapita GDP, taxes, Value added in agriculture, industry and services, taxes, Gini index, average salary and poverty rates. Goldsmith et al. 2019 recommend that the changes in the outcomes from the initial period to the start period of the analysis is used.

Table II.17: Correlations in Initial Period Covariates and Migration Decisions in the Subsequent Decade

	(1) Diff(per Capita GDP)	(2) Diff (Value Added -VA)	(3) Diff (VA agriculture)	(4) Diff (VA services)	(5) Diff (VA industry)	(6) Log(Exports) in 2010	(7) Log(Imports) in 2010
Migrants	0.00578	0.00600	0.0753	-0.0606	0.00912	6.76e-08	4.86e-08
	(0.00548)	(0.00548)	(0.0527)	(0.0372)	(0.00933)	(5.05e-08)	(3.61e-08)
Constant	1.202***	1.187***	1.692***	-0.279***	0.982***	0.000000450***	0.000000161***
	(0.000314)	(0.000315)	(0.00301)	(0.00213)	(0.000535)	(2.90e-09)	(2.07e-09)
Observations	44064	44064	44008	44048	44064	44528	44528
Adjusted R sq	0.2755	0.2745	0.5462	0.6072	0.2677	0.2416	0.2788
Microregion and Year F.E	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This figure shows correlation between differences in the initial period covariates (from 2010 to 2000) with the migration location decisions in the subsequent decade. The covariates include log differences in percapita GDP, Value added in agriculture, industry and services as well as exports and imports. Goldsmith et al. 2018 recommend that the changes in the outcomes in the initial period.

Table II.18: Pre-trends correlations

	(1) Total Exports in USD	(2) Total Imports in USD	(3) Differentiated Exports in USD	(4) Homogeneous Exports in USD	
1995 IV X T=2001	0.00000137	0.00000299	-0.000000864	-0.000000759	
	(0.00000416)	(0.00000396)	(0.00000413)	(0.00000392)	
1995 IV X T=2002	0.00000135	-0.00000116	0.00000286	-0.000000587	
	(0.00000436)	(0.00000387)	(0.00000426)	(0.00000388)	
1995 IV X T=2003	-0.00000349	0.00000205	0.00000310	-0.000000759	
	(0.00000447)	(0.00000384)	(0.00000416)	(0.00000382)	
1995 IV X T=2004	0.00000437	0.00000318	0.00000669	-0.000000930	
	(0.00000457)	(0.00000416)	(0.00000441)	(0.00000410)	
1995 IV X T=2005	0.00000663	0.00000801*	0.00000632	0.00000117	
	(0.00000430)	(0.00000437)	(0.00000427)	(0.00000405)	
1995 IV X T=2006	0.00000579	0.00000148	0.00000384	0.00000140	
	(0.00000470)	(0.00000445)	(0.00000454)	(0.00000433)	
1995 IV X T=2007	0.00000363	0.000000502	-0.000000877	0.00000148	
	(0.00000444)	(0.00000453)	(0.00000430)	(0.00000454)	
1995 IV X T=2008	0.000000762	0.00000158	0.00000105	0.00000314	
	(0.00000452)	(0.00000426)	(0.00000436)	(0.00000447)	
1995 IV X T=2009	0.0000102**	0.00000885**	0.00000416	0.00000551	
	(0.00000450)	(0.00000401)	(0.00000424)	(0.00000451)	
1995 IV X T=2010	0.0000236*	0.0000304**	0.0000106	0.0000141	
	(0.0000133)	(0.0000130)	(0.0000131)	(0.0000129)	
1995 IV X T=2011	0.0000122	0.00000258	0.00000330	0.0000113	
	(0.0000122)	(0.0000113)	(0.0000112)	(0.0000103)	
1995 IV X T=2012	0.00000891	0.0000125	0.00000899	0.0000161*	
	(0.0000101)	(0.0000101)	(0.00000970)	(0.00000968)	
1995 IV X T=2013	0.00000592	0.00000322	0.00000328	0.000000869	
	(0.00000743)	(0.00000646)	(0.00000722)	(0.00000760)	
1995 IV X T=2014	-0.00000116	0.0000172**	0.00000385	0.00000189	
	(0.00000774)	(0.00000690)	(0.00000722)	(0.00000714)	
1995 IV X T=2015	-0.00000917	-0.00000177	-0.00000648	0.00000361	
	(0.00000947)	(0.00000887)	(0.00000837)	(0.00000892)	
1995 IV X T=2016	0.0000125	0.00000343	0.000000199	0.0000139*	
	(0.00000774)	(0.00000800)	(0.00000733)	(0.00000836)	
1995 IV X T=2017	0.00000989	0.00000761	0.00000251	0.00000843	
	(0.00000650)	(0.00000665)	(0.00000606)	(0.00000714)	
Observations	29377	29377	29377	29377	29377
Year FE	Yes	Yes	Yes	Yes	Yes
UF F.E	Yes	Yes	Yes	Yes	Yes

Notes: This table shows the pre-trend and post-trend correlations of the IV interacted with time dummies with the outcome variables.

[50] I also perform the same robustness check at different levels of fixed effects (microregion and mesoregion) and find no correlation between the language shares and trade outcomes in the years leading up to the period of analysis.

II.A.1 Disentangling Short and Long term Impacts: Correlation Between Migration Cohorts

Recent critiques of the shift share IV indicate that the second stage point estimate might be conflating both short term and long term response. It would have been ideal if I can estimate the the short term and long term response based on the critique of Bartik Shift-Share IV based on Jaegger et al., 2018. To be able to disentangle the short and long term response, there should be independent variation across the two periods in the country of origin of immigrants. However, this is a very demanding criteria in my context. Jaegger et al. (2018) study a specific period between 1970 and 1980 where a policy change and other shocks in the United States dramatically changed the immigrant composition.

In my setting I find that the correlation between the instrument and its lags ranges between 0.89 to 0.98. Since the Bartik instruments are highly serially correlated, there will not be enough variation to identify the short and long run dynamics (Mckenzie, 2018; Jaegger, 2018). And second, I present the pair-wise correlations across language migrant cohorts for the years under study, 2010 to 2017. I also show the pair-wise correlation between the language cohorts across time because the shift part of my Bartik instrument relies on the inflows per year per language group and not country of origin. I report the results in tables II.19 and II.20 which show a strong correlation across cohorts reflecting that within this period of immigration into Brazil, there was a certain stability in the language groups arriving into the country[51]. I also divide the cohorts between those that arrived before 2014 and those that arrived after, and find that there is a correlation of 0.83. And while the multiple instrumentation approach proposed by Jaegger et al. (2018) is important, the high correlation across cohorts per year prevents me from estimating the short and long term responses due to high collinearity.

[51]Between 2010 and 2017, there was more of a variation in the country of origin composition of immigrants entering Brazil, with pair-wise yearly correlation ranging between 0.5 and 0.49. When looking at two periods, one before 2014 and one after, I find the correlation across these two cohorts to be 0.75. However, I am not estimating the short and long run impacts because my Instrument relies on the yearly "shock" of language groups and not country of origin, and in this case, I find a high correlation in the IV and its lags which will not allow for a separate identification of the short and long run responses

Table II.19: Correlations Across Cohorts for All Language groups

	cohort 2010	cohort 2011	cohort 2012	cohort 2013	cohort 2014	cohort 2015	cohort 2016	cohort 2017
cohort 2010	1							
cohort 2011	0.996***	1						
cohort 2012	0.989***	0.995***	1					
cohort 2013	0.952***	0.969***	0.980***	1				
cohort 2014	0.914***	0.939***	0.954***	0.993***	1			
cohort 2015	0.873***	0.904***	0.924***	0.977***	0.995***	1		
cohort 2016	0.876***	0.906***	0.923***	0.979***	0.995***	0.997***	1	
cohort 2017	0.937***	0.952***	0.959***	0.988***	0.979***	0.962***	0.972***	1
Observations	129							

Note: This table shows the year pair-wise cohort correlations across language groups. The analysis is based on the SINCRE Migration dataset. *** p<0.01, ** p<0.05, * p<0.1.

Table II.20: Correlations Across Cohorts for Top 14 Language Groups

	cohort 2010	cohort 2011	cohort 2012	cohort 2013	cohort 2014	cohort 2015	cohort 2016	cohort 2017
cohort 2010	1							
cohort 2011	0.995***	1						
cohort 2012	0.983***	0.992***	1					
cohort 2013	0.931***	0.956***	0.972***	1				
cohort 2014	0.872***	0.908***	0.930***	0.990***	1			
cohort 2015	0.809***	0.853***	0.883***	0.966***	0.993***	1		
cohort 2016	0.816***	0.858***	0.884***	0.969***	0.993***	0.999***	1	
cohort 2017	0.922***	0.946***	0.956***	0.992***	0.982***	0.959***	0.966***	1
Observations	14							

Note: This table shows the year pair-wise cohort correlations across the top 14 language groups. The analysis is based on the SINCRE Migration dataset. *** p<0.01, ** p<0.05, * p<0.1.

II.A.2 Shift-Share IV Based on Country of Origin

Table II.21 shows the results from the first stage 2SLS IV estimation where the shift share IV is constructed based on the country of origin commonality between newly arriving immigrants and historical locations of migrant settlements. The IV shift share is calculated similarly to the IV used in this paper but it uses the commonality of the country of origin as opposed to the commonality of the language spoken. In particular, the country of origin IV is calculated as follows:

$$L_{j,i}^{country} = (\frac{Mig_{j,i}}{\sum_j Mig_{j,i}})_{t=0} \tag{II.A.24}$$

In this case $Mig_{j,i}$ is the number of migrants from country j in location i at time 0 (i.e. year 2000). In order to get our instrument $Z_{i,t}^{country}$ for each municipality, I sum:

$$Z_{i,t}^{country} = \sum_{j=1}^{J} L_{j,i}^{country} \times \Delta Mig_{j,t} \tag{II.A.25}$$

Where $\Delta Mig_{j,t}$ is the annual inflow of migrants from a certain country j to Brazil in a given year.

Table II.21: First Stage Estimates for All Migrants Based on Country of Origin

	(1) Migrants	(2) Migrants	(3) Migrants	(4) Migrants	(5) Migrants	(6) Migrants
IV First Stage						
IV	0.000241**	0.000245***	0.000241***	0.000160*	0.000166*	0.000205**
	(0.000110)	(0.0000715)	(0.0000715)	(0.0000924)	(0.0000929)	(0.0000993)
Observations	44543	44543	44535	44560	44543	44543
Kleibergen-Paap F-stat	4.8429	11.7609	11.4136	3.0125	3.1844	4.2762
Cragg-Donald F Stat	114.9268	114.0260	106.7360	20.4683	21.6371	31.2016
Municipality Controls	Yes	Yes	Yes	No	Yes	Yes
Year FE	Yes	Yes	No	No	No	Yes
Fixed Effect	Microregion	Mesoregion	Mesoregion X Year	Yes	Yes	Yes

Notes: The table shows first stage 2SLS results for an a shift share strategy based on the country of origin historic locations of immigrants. Data sources: IBGE data for control variables, COMEX for trade data and SINCRE for migration data. Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p <0.1.

II.B Supplementary Material: Descriptive Statistics

The following section presents additional descriptive statistics on immigration into Brazil within the period of analysis (2010-2017) as well as on the control and outcome variables used in the analysis.

Table II.22: Immigrants by Country of Origin: Total and Growth Rates (%)

	Total	**Percentage**	**Percentage**
	Total 2010-2017	Growth from 2000 (%)	Average Annual Growth 2000 2010 (%)
Country of Origin			
HTI	99651	187920.8%	315.1%
BOL	70822	1151.7%	1.9%
COL	45808	3852.4%	29.0%
ARG	38205	841.2%	6.3%
USA	38001	654.1%	-3.8%
CHN	33516	643.6%	5.2%
PRT	26437	312.6%	0.8%
PER	25694	795.9%	9.4%
CUB	25447	4257.4%	233.9%
FRA	23637	994.3%	7.4%
PRY	23455	970.0%	1.2%
VEN	23375	3658.0%	65.8%
ITA	22780	823.8%	4.2%
ESP	21131	881.9%	7.1%
DEU	19808	710.8%	-0.9%
PHL	19466	8218.8%	-1.7%
URY	17743	355.2%	12.2%
IND	16261	7063.4%	8.2%
JPN	14834	590.0%	3.4%
MEX	12652	2305.3%	4.1%
GBR	12222	994.2%	-5.6%
KOR	12063	929.3%	14.0%
CHL	10258	385.7%	9.4%
AGO	9884	438.6%	3.3%
NLD	6063	1016.6%	4.4%
NOR	5722	1879.9%	-10.4%
ECU	5662	1872.8%	20.4%
SEN	5348	8667.2%	83.2%
IDN	5209	4768.2%	-9.8%
POL	4825	2016.2%	5.5%

Notes: This table shows the total number of immigrants by country of origin for the top 30 countries that arrived to Brazil between 2010-2017. The table also reports the growth rate for those countries from their concentration pre 2000. I also report the average annual growth from 2010 to 2017. Data sources: SINCRE for migration data.

Table II.23: Migrant Characteristics by Nationality Across All Years

	Migrants							
	All Other		Mercosur		OECD		Total	
	Total	%	Total	%	Total	%	Total	%
Sex								
Female	84,809	31.0%	99,989	42.1%	84,754	30.1%	269,552	34.0%
Male	189,060	69.0%	137,786	57.9%	196,778	69.9%	523,624	66.0%
Total	273,869	100.0%	237,775	100.0%	281,532	100.0%	793,176	100.0%
Skill Level of Profession								
Low Skill	9,309	3.4%	5,622	2.4%	6,632	2.4%	21,563	2.7%
High Skill	15,648	5.7%	57,277	24.1%	13,604	4.8%	86,529	10.9%
Student	27,183	9.9%	34,855	14.7%	55,900	19.9%	117,938	14.9%
No Occupation	16,416	6.0%	8,631	3.6%	7,078	2.5%	32,125	4.1%
Total	273,869	100.0%	237,775	100.0%	281,532	100.0%	793,176	100.0%
Age at Arrival								
0-15	10,510	3.8%	15,660	6.6%	15,850	5.6%	42,020	5.3%
15-25	39,167	14.3%	70,962	29.8%	59,673	21.2%	169,802	21.4%
25-40	136,803	50.0%	90,730	38.2%	100,742	35.8%	328,275	41.4%
40-65	51,313	18.7%	31,682	13.3%	81,622	29.0%	164,617	20.8%
65+	36,076	13.2%	28,741	12.1%	23,645	8.4%	88,462	11.2%
Total	273,869	100.0%	237,775	100.0%	281,532	100.0%	793,176	100.0%
Visa Type								
Border	14	0.0%	4,376	1.8%	58	0.0%	4,448	0.6%
Other	4,301	1.6%	445	0.2%	21	0.0%	4,767	0.6%
Permanent	112,728	41.2%	74,306	31.3%	95,880	34.1%	282,914	35.7%
Provisional	417	0.2%	52	0.0%	49	0.0%	518	0.1%
Temporary	128,670	47.0%	151,695	63.8%	180,730	64.2%	461,095	58.1%
Resident	27,729	10.1%	6,897	2.9%	4,791	1.7%	39,417	5.0%
Total	273,859	100.0%	237,771	100.0%	281,529	100.0%	793,159	100.0%

Notes: This table represents the characteristics of migrants in Brazil from 2010 to 2017. I present the characteristics on sex, skill level, age, civil status and visa type for immigrants from all countries, OECD, and Mercosur countries. Source: author's analysis of the SINCRE data.

Table II.24: Description of Dependent and Independent Variables

Variable	Description
Total Imports	Total Imports (USD or net KG) per municipality or per country-municipality.
Total Exports	Total Exports (USD or net KG) per municipality or per country-municipality.
Differentiated Goods	Differentiated Goods per municipality as per Rauch (1999) classification.
Homogeneous Goods	Homogeneous Goods per municipality as per Rauch (1999) classification. The category combines both homogeneous and reference goods.
Highly (High) Differentiated Goods	Elasticity of Substitution $\sigma \leq 2$ based on Broda and Weinstein (2006).
Moderately (Mid) Differentiated Goods	Elasticity of Substitution $2 < \sigma \leq 3.5$ based on Broda and Weinstein (2006).
Low (Low) Differentiated Goods	Elasticity of Substitution $\sigma > 3.5$ based on Broda and Weinstein (2006).

Notes: This table describes the main outcome variables used in the analysis. The analysis relies on traded goods based on Rauch classification of goods as well as subdivided based on σ elasticity of substitution based on Broda and Weinstein (2006)

II.C Supplementary Material: Robustness Checks and Different Specifications for Results

II.C.1 General Economy Set-up

What follows describes the key components[52] of Tai (2009)'s monopolistic competition model with firm heterogeneity and a multi-sector economy. The model considers a utility function with constant elasticity of substitution (CES) preferences. Assuming there are $H+1$ sectors with sector specific elasticities of substitution[53], σ_h. H is the number of differentiated sectors, whereby the homogeneous good is the numeraire. Each sector has Ω_h varieties and a consumer consumes q_o units of the good o and $q_h(w)$ units of each variety w of good h. There exists a preference parameter, a_{ij}, that affects demand for foreign goods between Brazilian municipality i and country j. Then utility can be expressed as follows:

$$U \equiv q_o^{\mu_o} \prod_{h=1}^{H} \left(\int_{\Omega_h} \left(a_{ij} q_h(w)\right)^{\frac{\sigma_h - 1}{\sigma_h}} dw \right)^{\frac{\sigma_h}{\sigma_h - 1}\mu_h}, \tag{II.C.26}$$

with $\mu_o + \sum_{h=1}^{H} \mu_h = 1$ and $\sigma_h > 1$.

The model assumes that firm productivity (ϕ) varies by sector and follows a Pareto distribution $dG_h(\phi) = \lambda_h \phi^{(-\lambda_h - 1)}$, where $\lambda \geq \sigma - 1$. The cost of producing q units and selling them in country j is represented as follows:

$$c_{ij}^h(\phi) = \frac{w_i \tau_{ij}^h}{\phi} q + f_{ij}^h, \tag{II.C.27}$$

where w_i represents the wage. Fixed costs are denoted by f_{ij}^h and variable costs (iceberg) are denoted by τ_{ij}^h.

From equations (1) and (2), a demand equation can be derived. Let Y_j denote the sum of income in country j with dividends from firms from j located worldwide.

[52] This section includes the key equations from Tai (2009) model to provide a sketch of the model to guide the discussion. I refer readers to Tai (2009) and Chaney (2008) for details on the model, proofs and derivations.

[53] This represents the product elasticity of substitution provided by Broda and Weinstein (2006), which is used for the empirical analysis.

$$x_{ij}^h(\phi) = p_{ij}^h(\phi) q_{ij}^h(\phi) = \mu_h Y_j \left(\frac{p_{ij}^h(\phi)}{a_{ij} p_j^h} \right)^{1-\sigma_h}, \tag{II.C.28}$$

where P_j^h is the price index in sector h for country j and p_{ij} is the selling price.

Studying cost mechanisms through which migrants from country j impact trade in municipality i is important because of the former's ability to obtain preferential access to market opportunities and the contacts that the migrant might have in their home country.

The threshold at which a firm is able to export is determined by the following equation, for a given productivity level ϕ. The productivity cut-off is the zero profits level, which has the following expression:

$$\bar{\phi}_{ijh} = \lambda_1^h \left(\frac{f_{ij}}{Y_j} \right)^{\frac{1}{\sigma_h - 1}} \frac{w_i \tau_{ij}}{P_j}, \tag{II.C.29}$$

where λ_h^1 is a sector specific constant[54].

The price index is defined as

$$P_j^{1-\sigma} = \sum_{k=1}^{N} w_k L_k \int_{\bar{\phi}_{kj}}^{\infty} \left(\frac{\sigma_h}{\sigma_h - 1} \times \frac{w_k \tau_{kj}}{\phi} \right)^{(1-\sigma_h)} dG(\phi), \tag{II.C.30}$$

with the determining threshold $P_j^h = \lambda_2^h \times Y_j^{\frac{1}{h} - \frac{1}{\sigma_h - 1}} \times \theta_j^h$. And, $(\theta_j^h)^{-\lambda_h} \equiv \sum_{k=1}^{N} (\frac{Y_k}{Y}) \times (w_k \tau_{kj})^{-\gamma} \times f_{kj}^{-(\frac{\lambda}{\sigma - 1} - 1)}$.

Y is world output, L_k is the labour income of country k, and λ_2^h is a constant by sector. Then with this price index defined, the demand equation for firm-level exports can be rewritten as:

$$x_{ij}(\phi | \phi \geq \bar{\phi}_{ij}) = \lambda_3^h \times \left(\frac{Y_j}{Y} \right)^{\left(\frac{\sigma_h - 1}{\lambda_h} \right)} \times a_{ij}^{\sigma_h - 1} \times \left(\frac{\theta_j}{w_i \tau_{ij}} \right)^{\sigma_h - 1} \times \phi^{\sigma_h - 1} \tag{II.C.31}$$

Then aggregate exports can be expressed as follows:

$$X_{ij}^h = \mu_h \times \frac{Y_i Y_j}{Y} \times a_{ij}^{\sigma_h - 1} \times \frac{w_i \tau_{ij}}{\theta_j} \times (f_{ij}^h)^{-(\frac{\lambda_h}{\sigma_h - 1} - 1)} \tag{II.C.32}$$

[54] I refer readers to Tai (2009) p. 232 for a description of this variable.

The underpinnings of the main equation used in section 4.2 are as presented in Tai (2009) as follows:

The parameters of preferences (a_{ijt}), fixed costs (f_{ijt}) and variable costs (τ_{ij}) are captured by:

$$a_{ijt} = mig_{ijt}^{\beta} \times distance_{ij} \times \exp^{language_{ij}+colony_{ij}+border_{ij}} \tag{II.C.33}$$

$$f_{ijt} = mig_{ijt}^{(1-\beta)} \times \exp^{language_{ij}+colony_{ij}} \tag{II.C.34}$$

$$\tau_{ij} = distance_{ij} \times \exp^{language_{ij}+colony_{ij}+border_{ij}} \tag{II.C.35}$$

The estimated equation is then:

$$lnX_{i,j,t} = \beta_1 ln(gdp_{it} \times gdp_{jt}) + \beta_2 ln(pop_{it} \times pop_{jt}) + \gamma_{j,i} + \beta ln(mig_{i,j,t}) + \epsilon_{i,j,t} \tag{II.C.36}$$

In order to estimate equation (II.C.36), I estimated the following equation using a rich set of fixed effects:

$$X_{i,j,t} = \upsilon_{i,t} + \gamma_{j,i} + \phi_{j,t} + \beta ln(mig_{i,j,t}) + \nu\chi_{i,t} + \epsilon_{i,j,t} \tag{II.C.37}$$

The following table represents the results from equation that links the model assumptions to the data.

Table II.25: Exports and Imports disaggregated by σ Interaction

	$\sigma \leq$ **75$^{\text{th}}$ percentile**	$\sigma \leq$ **95$^{\text{th}}$ percentile**
	Trade	Trade
Migrants x σ	0.606***	0.372***
	(0.179)	(0.0891)
Migrants x σ^2	-0.346**	-0.161***
	(0.137)	(0.0411)
Migrants x σ^3	0.0537**	0.0193***
	(0.0256)	(0.00550)
Municipality-Year F.E	Yes	Yes
Country-Municipality F.E	Yes	Yes
Country-Year F.E	Yes	Yes

Notes: This table presents the results from a PPML estimation of equation (4.6). The dependent variable is of the form $log(x+1)$ and the explanatory variable of immigration is the ratio of immigrants to the population per municipality. Note that this value is equivalent to the logarithmic form of the ratio of immigration to the population plus one. For the interpretation of a change in one percentage point, the ratio is multiplied by a 100. σ represents the elasticity of substitution of each exported or imported good. Elasticities of substitution σ are from Broda and Weinstein (2006). Standard errors are clustered at the municipality level. Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1.

II.C.2 OLS and PPML Estimations: Varying the Levels of Fixed Effects

It can be argued that these levels of fixed effects can be very restrictive and can reduce variation, driving results towards zero. Therefore, as robustness checks, I estimate (II.C.36) using municipality and country level controls as well as fixed effects as the state level. So, instead of using municipality-year and municipality-country fixed effects, I use state-year and state-country fixed effects. Notation remains as explained before in section 4.2, however, here s represents state. I estimate equation (II.C.38) and I present the results in the following tables.

$$X_{i,j,t} = \upsilon_{s,t} + \gamma_{j,s} + \phi_{j,t} + \beta ln(mig_{i,j,t}) + \nu\chi_{i,t} + \beta_1 ln(gdp_{it} \times gdp_{jt}) \qquad \text{(II.C.38)}$$
$$+ \beta_2 ln(pop_{it} \times pop_{jt}) + ln(Distance) + \epsilon_{i,j,t}$$

Table II.26: Effect of Immigration on Trade based on Rauch Classification (USD)

	(1)	(2)	(3)	(4)	(5)	(6)
	Total Exports	Differentiated Exports	Homogeneous Exports	Total Exports	Differentiated Exports	Homogeneous Exports
OLS and PPML analysis						
Migrants	0.148	0.390	0.734***	0.266***	0.187*	0.321***
	(0.200)	(0.246)	(0.255)	(0.0716)	(0.0972)	(0.0846)
I(Zero Mig)	-1.305***	-1.460***	-1.192***	-0.314**	-0.307**	-0.216
	(0.187)	(0.168)	(0.226)	(0.124)	(0.122)	(0.221)
log(GDP Masses)	0.454***	0.766***	0.785***	0.660***	0.711***	0.739***
	(0.0565)	(0.199)	(0.0580)	(0.0619)	(0.0423)	(0.0739)
log(Distance)	-8.851**	-3.602	-4.838	1.600	1.458	4.244
	(3.987)	(3.521)	(4.080)	(3.784)	(6.680)	(2.764)
	(1)	(2)	(3)	(4)	(5)	(6)
	Total Imports	Differentiated Imports	Homogeneous Imports	Total Imports	Differentiated Imports	Homogeneous Imports
Migrants	0.889**	1.001**	0.796**	0.450***	0.456***	0.279***
	(0.355)	(0.393)	(0.330)	(0.114)	(0.105)	(0.0812)
I(Zero Mig)	-1.441***	-1.723***	-1.767***	-0.175	-0.337***	0.111
	(0.138)	(0.108)	(0.134)	(0.124)	(0.0780)	(0.185)
log(GDP Masses)	1.608***	1.378***	1.264***	0.989***	1.049***	0.975***
	(0.0453)	(0.0695)	(0.0433)	(0.0487)	(0.0360)	(0.0665)
log(Distance)	7.670***	6.524***	4.849***	4.222	-1.433	-1.367
	(2.026)	(1.936)	(1.459)	(4.349)	(3.999)	(4.199)
Sample	All	All	All	All	All	All
Estimation	OLS	OLS	OLS	PPML	PPML	PPML
R squared	0.4759	0.4629	0.3365	0.6906	0.7628	0.5830
Municipality Controls	Yes	Yes	Yes	Yes	Yes	Yes
State-Year F.E	Yes	Yes	Yes	Yes	Yes	Yes
Country-State F.E	Yes	Yes	Yes	Yes	Yes	Yes
Country-Year F.E	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table shows the OLS and PPML estimates from equation (II.C.38). The dependent variable is of the form $log(x+1)$ and the explanatory variable of immigration is the ratio of immigrants to the population per municipality (in %). I(Zero Mun) is a dummy variable that indicates 1 if a municipality has no immigration. Estimations include municipality time varying characteristics, state-year, country-year and country-year F.E. The R^2 in the PPML estimations are Psuedo R^2, while those in the OLS estimations are the adjusted R^2. Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1.

Table II.27: **Extensive Margin**: Effect of Immigration on Trade based on Rauch Classification (USD)

	(1) Total Exports	(2) Differentiated Exports	(3) Homogeneous Exports	(4) Total Exports	(5) Differentiated Exports	(6) Homogeneous Exports
OLS and PPML analysis						
Migrants	0.561*	0.580**	0.421	0.205***	0.108	0.278***
	(0.325)	(0.274)	(0.332)	(0.0754)	(0.128)	(0.0943)
I(Zero Mig)	-1.459***	-1.291***	-0.874***	-0.449***	-0.407***	-0.348
	(0.130)	(0.124)	(0.146)	(0.122)	(0.113)	(0.218)
log(GDP Masses)	0.528***	0.480***	0.294***	0.416***	0.522***	0.514***
	(0.0785)	(0.109)	(0.0381)	(0.0796)	(0.0641)	(0.0935)
log(Distance)	-4.149***	-3.175**	-2.614**	-0.744	-1.989	1.102
	(1.415)	(1.468)	(1.052)	(4.088)	(5.293)	(3.820)
	(1) Total Imports	(2) Differentiated Imports	(3) Homogeneous Imports	(4) Total Imports	(5) Differentiated Imports	(6) Homogeneous Imports
Migrants	0.484	0.348	0.330	0.297***	0.424***	0.131**
	(0.330)	(0.309)	(0.260)	(0.0486)	(0.0838)	(0.0648)
I(Zero Mig)	-1.556***	-1.525***	-1.350***	-0.377***	-0.445***	-0.150
	(0.104)	(0.0577)	(0.0751)	(0.0780)	(0.122)	(0.139)
log(GDP Masses)	1.133***	0.929***	0.633***	0.759***	0.738***	0.781***
	(0.0552)	(0.0571)	(0.0420)	(0.0383)	(0.0469)	(0.0706)
log(Distance)	4.765***	4.540***	2.507**	0.508	-1.456	0.332
	(1.480)	(1.345)	(0.997)	(2.107)	(3.202)	(1.992)
Sample	All	All	All	All	All	All
Estimation	OLS	OLS	OLS	PPML	PPML	PPML
R squared	0.3685	0.3563	0.2170	0.4850	0.5054	0.4628
Municipality Controls	Yes	Yes	Yes	Yes	Yes	Yes
State-Year F.E	Yes	Yes	Yes	Yes	Yes	Yes
Country-State F.E	Yes	Yes	Yes	Yes	Yes	Yes
Country-Year F.E	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table shows the OLS and PPML estimates from equation (II.C.38) at the product extensive margin. The dependent variable is of the form $log(x+1)$ and the explanatory variable of immigration is the ratio of immigrants to the population per municipality (in %). I(Zero Mun) is a dummy variable that indicates 1 if a municipality has no immigration. Estimations include municipality time varying characteristics, state-year, country-year and country-year F.E. The R^2 in the PPML estimations are Psuedo R^2, while those in the OLS estimations are the adjusted R^2. Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1.

Table II.28: Effect of Immigration on Trade based on σ Elasticity of Substitution (USD)

	(1) log(Exports Low Diff.)	(2) log(Exports Mid Diff.)	(3) log(Exports High Diff.)	(4) Exports Low Diff.	(5) Exports Mid Diff.	(6) Exports High Diff.
OLS and PPML analysis						
Migrants	0.598	0.281	0.978***	0.309***	0.0571	0.320***
	(0.353)	(0.250)	(0.249)	(0.0852)	(0.352)	(0.0436)
I(Zero Mig)	-1.351***	-1.532***	-1.069***	-0.327*	-0.291***	0.0744
	(0.161)	(0.245)	(0.232)	(0.186)	(0.0945)	(0.185)
log(GDP Masses)	1.115***	0.571***	0.383***	0.780***	0.520***	0.585***
	(0.0734)	(0.102)	(0.0851)	(0.0605)	(0.0762)	(0.110)
log(Distance)	-3.429	-10.89**	-4.608***	3.803	-3.356	5.260
	(3.092)	(5.010)	(1.263)	(5.579)	(3.545)	(5.240)
	(1) log(Import Low Diff.)	(2) log(Import Mid Diff.)	(3) log(Import High Diff.)	(4) Import Low Diff.	(5) Import Mid Diff.	(6) Import High Diff.
Migrants	0.942***	0.952**	0.559	0.363***	0.521***	0.360***
	(0.313)	(0.346)	(0.572)	(0.0890)	(0.166)	(0.0905)
I(Zero Mig)	-1.770***	-1.749***	-1.781***	-0.0363	-0.303*	-0.276*
	(0.116)	(0.146)	(0.0740)	(0.109)	(0.177)	(0.145)
log(GDP Masses)	1.414***	1.471***	0.744***	1.030***	0.956***	0.894***
	(0.0556)	(0.0525)	(0.0727)	(0.0515)	(0.0575)	(0.0729)
log(Distance)	5.610***	6.857***	3.015**	-5.718	10.75*	11.94***
	(1.735)	(1.574)	(1.351)	(4.255)	(6.447)	(3.715)
Sample	All	All	All	All	All	All
Estimation	OLS	OLS	OLS	PPML	PPML	PPML
R squared	0.4307	0.4047	0.3008	0.6758	0.6669	0.6513
Municipality Controls	Yes	Yes	Yes	Yes	Yes	Yes
State-Year F.E	Yes	Yes	Yes	Yes	Yes	Yes
Country-State F.E	Yes	Yes	Yes	Yes	Yes	Yes
Country-Year F.E	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table shows the OLS and PPML estimates from equation (II.C.38). The dependent variable is of the form $log(x+1)$ and the explanatory variable of immigration is the ratio of immigrants to the population per municipality (in %). I(Zero Mun) is a dummy variable that indicates 1 if a municipality has no immigration. Estimations include municipality time varying characteristics, state-year, country-year and country-year F.E. The R^2 in the PPML estimations are Psuedo R^2, while those in the OLS estimations are the adjusted R^2. Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1.

Table II.29: **Extensive Margin:** Effect of Immigration on Trade based on σ Elasticity of Substitution (USD)

	(1)	(2)	(3)	(4)	(5)	(6)
	Exports Low Diff.	Exports Mid Diff.	Exports High Diff.	Exports Low Diff.	Exports Mid Diff.	Exports High Diff.
OLS and PPML analysis						
Migrants	0.650*	0.501*	0.626**	0.232**	0.182*	0.207***
	(0.380)	(0.283)	(0.239)	(0.0911)	(0.0986)	(0.0532)
I(Zero Mig)	-1.182***	-1.224***	-0.567***	-0.471***	-0.356**	-0.190
	(0.149)	(0.152)	(0.0999)	(0.135)	(0.164)	(0.141)
log(GDP Masses)	0.503***	0.370***	0.121***	0.590***	0.297***	0.234**
	(0.0641)	(0.0652)	(0.0329)	(0.0792)	(0.0867)	(0.101)
log(Distance)	-2.364**	-4.075**	-1.885***	0.160	-2.524	1.393
	(1.073)	(1.531)	(0.550)	(6.524)	(3.597)	(4.442)
	(1)	(2)	(3)	(4)	(5)	(6)
	Import Low Diff.	Import Mid Diff.	Import High Diff.	Import Low Diff.	Import Mid Diff.	Import High Diff.
Migrants	0.275	0.447	0.290	0.299***	0.332***	0.266***
	(0.288)	(0.273)	(0.275)	(0.0582)	(0.0502)	(0.0459)
I(Zero Mig)	-1.529***	-1.533***	-0.968***	-0.484***	-0.264**	-0.392***
	(0.0616)	(0.0734)	(0.0434)	(0.0848)	(0.103)	(0.122)
log(GDP Masses)	0.865***	0.886***	0.321***	0.758***	0.764***	0.713***
	(0.0500)	(0.0520)	(0.0313)	(0.0368)	(0.0459)	(0.0599)
log(Distance)	3.373***	4.117***	1.424**	-0.594	1.098	3.979
	(1.112)	(1.280)	(0.680)	(2.984)	(2.182)	(3.210)
Sample	All	All	All	All	All	All
Estimation	OLS	OLS	OLS	PPML	PPML	PPML
R squared	0.3140	0.3022	0.1680	0.5159	0.4398	0.4170
Municipality Controls	Yes	Yes	Yes	Yes	Yes	Yes
State-Year F.E	Yes	Yes	Yes	Yes	Yes	Yes
Country-State F.E	Yes	Yes	Yes	Yes	Yes	Yes
Country-Year F.E	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table shows the OLS and PPML estimates from equation (II.C.38) at the product extensive margin. The dependent variable is of the form $log(x + 1)$ and the explanatory variable of immigration is the ratio of immigrants to the population per municipality (in %). I(Zero Mun) is a dummy variable that indicates 1 if a municipality has no immigration. Estimations include municipality time varying characteristics, state-year, country-year and country-year F.E. The R^2 in the PPML estimations are Psuedo R^2, while those in the OLS estimations are the adjusted R^2. Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1.

I then proceed to estimate equation (II.C.38), but instead of using state-year fixed effects, I use municipality-year fixed effects.

Table II.30: Effect of Immigration on Trade based on Rauch Classification

	(1) Total Exports	(2) Differentiated Exports	(3) Homogeneous Exports	(4) Total Exports	(5) Differentiated Exports	(6) Homogeneous Exports
OLS and PPML analysis						
Migrants	0.991***	0.952***	1.107***	0.953	3.345***	0.923
	(0.339)	(0.322)	(0.419)	(0.896)	(0.962)	(0.610)
I(Zero Mig)	-1.265***	-1.318***	-1.423***	-0.263***	-0.197***	-0.191**
	(0.0646)	(0.0649)	(0.0811)	(0.0596)	(0.0631)	(0.0917)
log(Distance)	-4.808***	-1.744*	-5.709***	-1.104	-2.376	-1.096
	(1.178)	(0.980)	(1.198)	(2.098)	(3.955)	(2.242)
Country-Mun Border	-0.694*	-0.803*	-1.135***	-0.199	-0.895	0.183
	(0.386)	(0.425)	(0.380)	(0.440)	(0.700)	(0.451)
	(1) Total Imports	(2) Differentiated Imports	(3) Homogeneous Imports	(4) Total Imports	(5) Differentiated Imports	(6) Homogeneous Imports
Migrants	0.580**	0.827***	0.855**	4.061***	3.581**	3.424***
	(0.279)	(0.302)	(0.384)	(0.932)	(1.396)	(1.140)
I(Zero Mig)	-1.350***	-1.435***	-1.730***	-0.279***	-0.367***	-0.184
	(0.0569)	(0.0597)	(0.0742)	(0.0923)	(0.0697)	(0.113)
log(Distance)	5.700***	6.436***	4.837***	-5.674	-10.15**	-6.833**
	(0.995)	(0.829)	(0.853)	(4.012)	(4.631)	(3.221)
Country-Mun Border	0.785**	-0.0487	0.690***	0.543	0.140	-0.301
	(0.312)	(0.243)	(0.261)	(0.460)	(0.668)	(0.646)
Observations	385499	385499	385499	353502	332912	308603
Sample	All	All	All	All	All	All
Municipalities	4342	4342	4342	4342	4342	4342
Estimation	OLS	OLS	OLS	PPML	PPML	PPML
Mean Migrants	0.00492	0.00492	0.00492	0.00492	0.00492	0.00492
R squared	0.5820	0.5441	0.4271	0.8233	0.8643	0.8164
Municipality-Year F.E	Yes	Yes	Yes	Yes	Yes	Yes
Country-State F.E	Yes	Yes	Yes	Yes	Yes	Yes
Country-Year F.E	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table shows the OLS and PPML estimates from equation (II.C.38). The dependent variable is of the form $log(x+1)$ and the explanatory variable of immigration is the ratio of immigrants to the population per municipality (in %). I(Zero Mun) is a dummy variable that indicates 1 if a municipality has no immigration. Estimations include municipality-year, country-year and country-year F.E. The R^2 in the PPML estimations are Psuedo R^2, while those in the OLS estimations are the adjusted R^2. Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1.

Table II.31: **Extensive Margin:** Effect of Immigration on Trade on the Extensive margin (in USD) based on Rauch Classification

	(1)	(2)	(3)	(4)	(5)	(6)
	Total Exports	Differentiated Exports	Homogeneous Exports	Total Exports	Differentiated Exports	Homogeneous Exports
OLS and PPML analysis						
Migrants	0.825***	0.778***	0.514	0.204	2.177***	-0.378
	(0.286)	(0.263)	(0.332)	(0.662)	(0.829)	(0.840)
I(Zero Mig)	-1.345***	-1.189***	-0.891***	-0.323***	-0.324***	-0.236**
	(0.0625)	(0.0578)	(0.0581)	(0.0773)	(0.0914)	(0.117)
log(Distance)	-2.943***	-2.383***	-2.772***	-3.012**	-4.261*	-5.396**
	(0.626)	(0.515)	(0.457)	(1.402)	(2.265)	(2.723)
Country-Mun Border	-0.920***	-0.757***	-0.752***	-0.349	-1.593***	0.367
	(0.280)	(0.254)	(0.179)	(0.329)	(0.463)	(0.304)
	(1)	(2)	(3)	(4)	(5)	(6)
	Total Imports	Differentiated Imports	Homogeneous Imports	Total Imports	Differentiated Imports	Homogeneous Imports
Migrants	0.282	0.211	0.369	0.355	2.757**	1.115
	(0.350)	(0.316)	(0.267)	(0.232)	(1.072)	(0.821)
I(Zero Mig)	-1.367***	-1.287***	-1.274***	-0.482***	-0.429***	-0.392***
	(0.0544)	(0.0512)	(0.0601)	(0.0680)	(0.0851)	(0.0856)
log(Distance)	5.016***	5.171***	3.474***	-2.196	-3.819	-2.582
	(0.626)	(0.583)	(0.416)	(1.834)	(2.690)	(2.191)
Country-Mun Border	0.182	0.0249	0.246**	-0.429	-1.138*	-0.308
	(0.163)	(0.127)	(0.105)	(0.455)	(0.629)	(0.546)
Observations	385499	385499	385499	342863	320485	281481
Sample	All	All	All	All	All	All
Municipalities	4342	4342	4342	4342	4342	4342
Estimation	OLS	OLS	OLS	PPML	PPML	PPML
Mean Migrants	0.00492	0.00492	0.00492	0.00492	0.00492	0.00492
R squared	0.4365	0.4164	0.2587	0.6533	0.6857	0.6724
Municipality-Year F.E	Yes	Yes	Yes	Yes	Yes	Yes
Country-State F.E	Yes	Yes	Yes	Yes	Yes	Yes
Country-Year F.E	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table shows the OLS and PPML estimates from equation (II.C.38) at the product extensive margin. The dependent variable is of the form $log(x+1)$ and the explanatory variable of immigration is the ratio of immigrants to the population per municipality (in %). I(Zero Mun) is a dummy variable that indicates 1 if a municipality has no immigration. Estimations include municipality-year, state-year, country-year and country-year F.E. The R^2 in the PPML estimations are Psuedo R^2, while those in the OLS estimations are the adjusted R^2. Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1.

Table II.32: Effect of Immigration on Trade (in USD) based on σ Elasticity of Substitution

	(1) log(Exports Low Diff.)	(2) log(Exports Mid Diff.)	(3) log(Exports High Diff.)	(4) Exports Low Diff.	(5) Exports Mid Diff.	(6) Exports High Diff.
OLS and PPML analysis						
Migrants	1.041**	0.765**	1.238***	1.232	1.833	3.865***
	(0.408)	(0.311)	(0.415)	(0.816)	(1.179)	(0.811)
I(Zero Mig)	-1.315***	-1.490***	-1.145***	-0.254***	-0.209***	0.0476
	(0.0719)	(0.0756)	(0.0768)	(0.0616)	(0.0735)	(0.0844)
log(Distance)	-3.024**	-5.615***	-3.407***	-4.434	-0.979	-3.708
	(1.284)	(1.221)	(0.839)	(3.077)	(1.979)	(2.439)
Country-Mun Border	-1.119**	-1.157***	-1.084***	-0.262	-0.0394	-0.724
	(0.505)	(0.381)	(0.307)	(0.537)	(0.339)	(0.613)
	(1) log(Import Low Diff.)	(2) log(Import Mid Diff.)	(3) log(Import High Diff.)	(4) Import Low Diff.	(5) Import Mid Diff.	(6) Import High Diff.
Migrants	0.828***	0.922**	0.489	4.542***	3.715***	0.537
	(0.319)	(0.383)	(0.488)	(1.115)	(1.113)	(1.130)
I(Zero Mig)	-1.545***	-1.585***	-1.630***	-0.248**	-0.309***	-0.123**
	(0.0629)	(0.0644)	(0.0688)	(0.118)	(0.0711)	(0.0528)
log(Distance)	4.444***	6.694***	3.795***	-21.43**	5.570	10.39***
	(0.839)	(0.954)	(0.554)	(9.405)	(3.653)	(3.186)
Country-Mun Border	0.231	0.893***	0.657***	0.326	0.457	1.761***
	(0.265)	(0.297)	(0.187)	(0.584)	(0.606)	(0.371)
Observations	385499	385499	385499	333679	333917	257257
Sample	All	All	All	All	All	All
Municipalities	4342	4342	4342	4342	4342	4342
Estimation	OLS	OLS	OLS	PPML	PPML	PPML
Mean Migrants	0.00492	0.00492	0.00492	0.00492	0.00492	0.00492
R squared	0.5107	0.4880	0.3569	0.8146	0.8187	0.7936
Municipality-Year F.E	Yes	Yes	Yes	Yes	Yes	Yes
Country-State F.E	Yes	Yes	Yes	Yes	Yes	Yes
Country-Year F.E	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table shows the OLS and PPML estimates from equation (II.C.38). The dependent variable is of the form $log(x+1)$ and the explanatory variable of immigration is the ratio of immigrants to the population per municipality (in %). I(Zero Mun) is a dummy variable that indicates 1 if a municipality has no immigration. Estimations include municipality-year, country-year and country-year F.E. The R^2 in the PPML estimations are Psuedo R^2, while those in the OLS estimations are the adjusted R^2. Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1.

Table II.33: **Extensive Margin:** Effect of Immigration on Trade (in USD) based on σ Elasticity of Substitution

	(1) Exports Low Diff.	(2) Exports Mid Diff.	(3) Exports High Diff.	(4) Exports Low Diff.	(5) Exports Mid Diff.	(6) Exports High Diff.
OLS and PPML analysis						
Migrants	0.783**	0.674**	0.667***	-0.314	1.413	1.962**
	(0.309)	(0.277)	(0.241)	(0.876)	(1.055)	(0.892)
I(Zero Mig)	-1.080***	-1.163***	-0.563***	-0.306***	-0.227**	-0.248*
	(0.0614)	(0.0588)	(0.0411)	(0.100)	(0.0959)	(0.127)
log(Distance)	-2.212***	-2.743***	-1.546***	-7.807***	0.517	0.00745
	(0.564)	(0.532)	(0.294)	(2.441)	(1.377)	(2.452)
Country-Mun Border	-0.645**	-0.937***	-0.432***	-0.491	-0.351	0.0169
	(0.250)	(0.239)	(0.109)	(0.464)	(0.328)	(0.489)
	(1) Import Low Diff.	(2) Import Mid Diff.	(3) Import High Diff.	(4) Import Low Diff.	(5) Import Mid Diff.	(6) Import High Diff.
Migrants	0.123	0.433	0.257	0.234*	1.635	0.960
	(0.300)	(0.287)	(0.249)	(0.133)	(1.417)	(0.839)
I(Zero Mig)	-1.327***	-1.354***	-0.882***	-0.519***	-0.414***	-0.253***
	(0.0553)	(0.0554)	(0.0436)	(0.0893)	(0.0751)	(0.0971)
log(Distance)	3.562***	5.225***	2.045***	-3.204	-2.185	4.994*
	(0.490)	(0.560)	(0.250)	(2.714)	(2.207)	(2.847)
Country-Mun Border	0.110	0.180	0.166***	-0.303	-0.739	0.606*
	(0.123)	(0.127)	(0.0637)	(0.529)	(0.629)	(0.349)
Observations	385499	385499	385499	316522	318463	221697
Sample	All	All	All	All	All	All
Municipalities	4342	4342	4342	4342	4342	4342
Estimation	OLS	OLS	OLS	PPML	PPML	PPML
Mean Migrants	0.00492	0.00492	0.00492	0.00492	0.00492	0.00492
R squared	0.3659	0.3550	0.1892	0.6815	0.6417	0.6198
Municipality-Year F.E	Yes	Yes	Yes	Yes	Yes	Yes
Country-State F.E	Yes	Yes	Yes	Yes	Yes	Yes
Country-Year F.E	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table shows the OLS and PPML estimates from equation (II.C.38) at the product extensive margin. The dependent variable is of the form $log(x+1)$ and the explanatory variable of immigration is the ratio of immigrants to the population per municipality (in %). I(Zero Mun) is a dummy variable that indicates 1 if a municipality has no immigration. Estimations include municipality-year, state-year, country-year and country-year F.E. The R^2 in the PPML estimations are Psuedo R^2, while those in the OLS estimations are the adjusted R^2. Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1.

II.C.3 IV 2SLS Estimations at the Country-Municipality Levels

This section presents the results from the estimation of equation (5.21) on exports and imports at the country and municipality levels.

Table II.34: Based on Rauch Classification

	(1) log(Total Exports in USD)	(2) log(Differentiated Exports in USD)	(3) log(Homogeneous Exports in USD)	(4) log(Total Imports in USD)	(5) log(Differentiated Imports in USD)	(6) log(Homogeneous Imports in USD)
Migrants	1.865**	1.036	0.667	3.100***	0.311	1.781***
	(0.786)	(0.641)	(0.698)	(0.800)	(0.502)	(0.656)
Estimation	IV	IV	IV	IV	IV	IV
Country-Municipality F.E	Yes	Yes	Yes	Yes	Yes	Yes
Municipality-Year F.E	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The table shows 2SLS results for subcategories of exports and imports based on the elasticity of substitution σ. Higher elasticity of substitution means that the product is less differentiated and hence more homogeneous. The dependent variable is of the form $log(x+1)$ and the explanatory variable of immigration is the ratio of immigrants to the population per municipality (in %). Robust standard errors in parentheses clustered at the mesoregion level. The predicted migration variable is divided by population in each municipality. Data sources: IBGE data for control variables, COMEX for trade data and SINCRE for migration data. Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1.

Table II.35: Based on σ Elasticity of Substitution

	(1) log(Exports Low Diff.)	(2) log(Exports Mid Diff.)	(3) log(Exports High Diff.)	(4) log(Import Low Diff.)	(5) log(Import Mid Diff.)	(6) log(Import High Diff.)
Migrants	2.151***	0.120	1.260*	2.741***	1.879**	-0.426
	(0.750)	(0.703)	(0.688)	(0.677)	(0.744)	(0.409)
Estimation	IV	IV	IV	IV	IV	IV
Country-Municipality F.E	Yes	Yes	Yes	Yes	Yes	Yes
Municipality-Year F.E	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The table shows 2SLS results for subcategories of exports and imports based on the elasticity of substitution σ. Higher elasticity of substitution means that the product is less differentiated and hence more homogeneous. The dependent variable is of the form $log(x+1)$ and the explanatory variable of immigration is the ratio of immigrants to the population per municipality (in %). Robust standard errors in parentheses clustered at the mesoregion level. The predicted migration variable is divided by population in each municipality. Data sources: IBGE data for control variables, COMEX for trade data and SINCRE for migration data. Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1.

Table II.36: Based on Rauch Classification on the Extensive margin

	(1) log(Total Exports in USD)	(2) log(Differentiated Exports in USD)	(3) log(Homogeneous Exports in USD)	(4) log(Total Imports in USD)	(5) log(Differentiated Imports in USD)	(6) log(Homogeneous Imports in USD)
Migrants	1.314	0.367	0.543	2.403***	0.709*	1.046**
	(0.850)	(0.672)	(0.693)	(0.664)	(0.412)	(0.475)
Estimation	IV	IV	IV	IV	IV	IV
Country-Municipality F.E	Yes	Yes	Yes	Yes	Yes	Yes
Municipality-Year F.E	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The table shows 2SLS results for subcategories of exports and imports based on Rauch Classification of goods. Higher elasticity of substitution means that the product is less differentiated and hence more homogeneous. The dependent variable is of the form $log(x+1)$ and the explanatory variable of immigration is the ratio of immigrants to the population per municipality (in %). Robust standard errors in parentheses clustered at the mesoregion level. The predicted migration variable is divided by population in each municipality. Data sources: IBGE data for control variables, COMEX for trade data and SINCRE for migration data. Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1.

Table II.37: Based on σ Elasticity of Substitution on the Extensive margin

	(1) log(Exports Low Diff.)	(2) log(Exports Mid Diff.)	(3) log(Exports High Diff.)	(4) log(Import Low Diff.)	(5) log(Import Mid Diff.)	(6) log(Import High Diff.)
Migrants	0.679	-0.150	0.306	1.737***	1.336**	-0.162
	(0.677)	(0.724)	(0.534)	(0.546)	(0.541)	(0.313)
Estimation	IV	IV	IV	IV	IV	IV
Country-Municipality F.E	Yes	Yes	Yes	Yes	Yes	Yes
Municipality-Year F.E	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The table shows 2SLS results for subcategories of exports and imports based on the elasticity of substitution σ. Higher elasticity of substitution means that the product is less differentiated and hence more homogeneous. The dependent variable is of the form $log(x+1)$ and the explanatory variable of immigration is the ratio of immigrants to the population per municipality (in %). Robust standard errors in parentheses clustered at the mesoregion level. The predicted migration variable is divided by population in each municipality. Data sources: IBGE data for control variables, COMEX for trade data and SINCRE for migration data. Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1.

II.C.4 IV 2SLS Estimations at the Municipality Level Using Different Levels of Fixed Effects.

In what follows, I re-estimate equation (5.15) using different levels of fixed effects than the municipality level used in the estimations presented earlier. In particular, I use Federation unit-year fixed effects, and mesoregion-year fixed effects.

Table II.38: Total Exports and Imports

	(1) log(Total Exports)	(2) log(Total Exports)	(3) log(Total Imports)	(4) log(Total Imports)
Migrants	0.0265	1.940***	0.0352	2.935***
	(0.0252)	(0.695)	(0.0306)	(1.044)
Observations	44543	44543	44543	44543
Estimation	OLS	IV	OLS	IV
Mean Migrants	0.000633	0.000633	0.000633	0.000633
Doubling of Mean	0.00167	0.123	0.00223	0.186
Municipality Controls	Yes	Yes	Yes	Yes
UFxYear F.E	Yes	Yes	Yes	Yes

Notes: The table shows OLS and 2SLS results for total export and imports. The dependent variable is of the form $log(x+1)$ and the explanatory variable of immigration is the ratio of immigrants to the population per municipality (in %). Robust standard errors in parentheses clustered at the Federation Unit level. Data sources: IBGE data for control variables, COMEX for trade data and SINCRE for migration data. Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1.

Table II.39: Differentiated and Homogeneous Exports based on Rauch Classification

	(1) log(Differentiated Exports)	(2) log(Differentiated Exports)	(3) log(Homogeneous Exports)	(4) log(Homogeneous Exports)
Migrants	0.0159	1.059**	0.0305	1.020**
	(0.0161)	(0.500)	(0.0254)	(0.445)
Observations	44543	44543	44543	44543
Estimation	OLS	IV	OLS	IV
Mean Migrants	0.000633	0.000633	0.000633	0.000633
Doubling of Mean	0.00101	0.0670	0.00193	0.0646
Municipality Controls	Yes	Yes	Yes	Yes
UFxYear F.E	Yes	Yes	Yes	Yes

Notes: The table shows OLS and 2SLS results for subcategories of exports based on Rauch Classification of goods. The dependent variable is of the form $log(x+1)$ and the explanatory variable of immigration is the ratio of immigrants to the population per municipality (in %). Robust standard errors in parentheses clustered at the Federation Unit level. The predicted migration variable is divided by population in each municipality. Data sources: IBGE data for control variables, COMEX for trade data and SINCRE for migration data. Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1.

Table II.40: Differentiated and Homogeneous Imports based on Rauch Classification

	(1) log(Differentiated Imports)	(2) log(Differentiated Imports)	(3) log(Homogeneous Imports)	(4) log(Homogeneous Imports)
Migrants	0.0334	2.629***	0.0174	1.688**
	(0.0286)	(0.945)	(0.0187)	(0.723)
Observations	44543	44543	44543	44543
Estimation	OLS	IV	OLS	IV
Mean Migrants	0.000633	0.000633	0.000633	0.000633
Doubling of Mean	0.00211	0.166	0.00110	0.107
Municipality Controls	Yes	Yes	Yes	Yes
UFxYear F.E	Yes	Yes	Yes	Yes

Notes: The table shows OLS and 2SLS results for subcategories of imports based on Rauch Classification of goods. The dependent variable is of the form $log(x+1)$ and the explanatory variable of immigration is the ratio of immigrants to the population per municipality (in %). Robust standard errors in parentheses clustered at the Federation Unit level. The predicted migration variable is divided by population in each municipality. Data sources: IBGE data for control variables, COMEX for trade data and SINCRE for migration data. Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1.

Table II.41: Based on σ Elasticity of Substitution

	(1) log(Exports Low Diff.)	(2) log(Exports Low Diff.)	(3) log(Exports Mid Diff.)	(4) log(Exports Mid Diff.)	(5) log(Exports High Diff.)	(6) log(Exports High Diff.)
Migrants	0.0283	1.104**	0.00814	0.755	0.0227	0.386
	(0.0272)	(0.511)	(0.0142)	(0.460)	(0.0175)	(0.305)
Observations	44543	44543	44543	44543	44543	44543
Estimation	OLS	IV	OLS	IV	OLS	IV
Mean Migrants	0.000633	0.000633	0.000633	0.000633	0.000633	0.000633
Doubling of Mean	0.00179	0.0699	0.000515	0.0478	0.00144	0.0244
	(1) log(Import Low Diff.)	(2) log(Import Low Diff.)	(3) log(Import Mid Diff.)	(4) log(Import Mid Diff.)	(5) log(Import High Diff.)	(6) log(Import High Diff.)
Migrants	0.0311	2.474***	0.0298	2.484***	0.0208	1.489**
	(0.0272)	(0.913)	(0.0259)	(0.914)	(0.0193)	(0.582)
Observations	44543	44543	44543	44543	44543	44543
Estimation	OLS	IV	OLS	IV	OLS	IV
Mean Migrants	0.000633	0.000633	0.000633	0.000633	0.000633	0.000633
Doubling of Mean	0.00197	0.157	0.00188	0.157	0.00132	0.0943
Municipality Controls	Yes	Yes	Yes	Yes	Yes	Yes
UFxYear F.E	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The table shows OLS and 2SLS results for subcategories of exports imports based on σ elasticity of substitution. The dependent variable is of the form $log(x+1)$ and the explanatory variable of immigration is the ratio of immigrants to the population per municipality (in %). Robust standard errors in parentheses clustered at the Federation Unit level. The predicted migration variable is divided by population in each municipality. Data sources: IBGE data for control variables, COMEX for trade data and SINCRE for migration data. Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1.

Table II.42: Total Exports and Imports

	(1) log(Total Exports)	(2) log(Total Exports)	(3) log(Total Imports)	(4) log(Total Imports)
Migrants	0.0282	2.242***	0.0369	2.730***
	(0.0262)	(0.787)	(0.0318)	(0.967)
Observations	44543	44543	44543	44543
Estimation	OLS	IV	OLS	IV
Mean Migrants	0.000633	0.000633	0.000633	0.000633
Doubling of Mean	0.00178	0.142	0.00233	0.173
Municipality Controls	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Mesoregion x Year F.E	Yes	Yes	Yes	Yes

Notes: The table shows OLS and 2SLS results for total export and imports. The dependent variable is of the form $log(x+1)$ and the explanatory variable of immigration is the ratio of immigrants to the population per municipality (in %). Robust standard errors in parentheses clustered at the Mesoregion level. The predicted migration variable is divided by population in each municipality. Data sources: IBGE data for control variables, COMEX for trade data and SINCRE for migration data. Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1.

Table II.43: Differentiated and Homogeneous Exports based on Rauch Classification

	(1) log(Differentiated Exports)	(2) log(Differentiated Exports)	(3) log(Homogeneous Exports)	(4) log(Homogeneous Exports)
Migrants	0.0219	1.107**	0.0279	1.322**
	(0.0187)	(0.501)	(0.0245)	(0.527)
Observations	44543	44543	44543	44543
Estimation	OLS	IV	OLS	IV
Mean Migrants	0.000633	0.000633	0.000633	0.000633
Doubling of Mean	0.00138	0.0701	0.00176	0.0837
Municipality Controls	Yes	Yes	Yes	Yes
Mesoregion x Year F.E	Yes	Yes	Yes	Yes

Notes: The table shows OLS and 2SLS results for subcategories of exports based on Rauch Classification of goods. The dependent variable is of the form $log(x+1)$ and the explanatory variable of immigration is the ratio of immigrants to the population per municipality (in %). Robust standard errors in parentheses clustered at the mesoregion level. The predicted migration variable is divided by population in each municipality. Data sources: IBGE data for control variables, COMEX for trade data and SINCRE for migration data. Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1.

Table II.44: Differentiated and Homogeneous Imports based on Rauch Classification

	(1) log(Differentiated Imports)	(2) log(Differentiated Imports)	(3) log(Homogeneous Imports)	(4) log(Homogeneous Imports)
Migrants	0.0389	2.510***	0.0212	1.671**
	(0.0320)	(0.881)	(0.0207)	(0.674)
Observations	44543	44543	44543	44543
Estimation	OLS	IV	OLS	IV
Mean Migrants	0.000633	0.000633	0.000633	0.000633
Doubling of Mean	0.00246	0.159	0.00134	0.106
Municipality Controls	Yes	Yes	Yes	Yes
Mesoregion x Year F.E	Yes	Yes	Yes	Yes

Notes: The table shows OLS and 2SLS results for subcategories of imports based on Rauch Classification of goods. The dependent variable is of the form $log(x+1)$ and the explanatory variable of immigration is the ratio of immigrants to the population per municipality (in %). Robust standard errors in parentheses clustered at the mesoregion level. The predicted migration variable is divided by population in each municipality. Data sources: IBGE data for control variables, COMEX for trade data and SINCRE for migration data. Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1.

Table II.45: Based on σ Elasticity of Substitution

	(1) log(Exports Low Diff.)	(2) log(Exports Low Diff.)	(3) log(Exports Mid Diff.)	(4) log(Exports Mid Diff.)	(5) log(Exports High Diff.)	(6) log(Exports High Diff.)
Migrants	0.0308	1.449**	0.00990	0.822*	0.0238	0.402
	(0.0285)	(0.576)	(0.0136)	(0.471)	(0.0176)	(0.304)
Observations	44543	44543	44543	44543	44543	44543
Estimation	OLS	IV	OLS	IV	OLS	IV
Mean Migrants	0.000633	0.000633	0.000633	0.000633	0.000633	0.000633
Doubling of Mean	0.00195	0.0917	0.000627	0.0520	0.00150	0.0254
	(1) log(Import Low Diff.)	(2) log(Import Low Diff.)	(3) log(Import Mid Diff.)	(4) log(Import Mid Diff.)	(5) log(Import High Diff.)	(6) log(Import High Diff.)
Migrants	0.0338	2.322***	0.0320	2.275***	0.0217	1.295**
	(0.0290)	(0.850)	(0.0270)	(0.837)	(0.0198)	(0.498)
Observations	44543	44543	44543	44543	44543	44543
Estimation	OLS	IV	OLS	IV	OLS	IV
Mean Migrants	0.000633	0.000633	0.000633	0.000633	0.000633	0.000633
Doubling of Mean	0.00214	0.147	0.00203	0.144	0.00137	0.0820
Municipality Controls	Yes	Yes	Yes	Yes	Yes	Yes
Mesoregion x Year F.E	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Notes: The table shows OLS and 2SLS results for subcategories of exports imports based on σ elasticity of substitution. The dependent variable is of the form $log(x+1)$ and the explanatory variable of immigration is the ratio of immigrants to the population per municipality (in %). Robust standard errors in parentheses clustered at the mesoregion level. The predicted migration variable is divided by population in each municipality. Data sources: IBGE data for control variables, COMEX for trade data and SINCRE for migration data. Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1.

II.C.5 Robustness Checks of 2SLS Estimations for Different Sample Composition

In this section, I present multiple robustness checks for the IV 2SLS estimations at the municipality level by varying the sample composition or the way the control variable are included. I present the results for the following samples of municipalities: 1- municipalities but excluding outlier municipalities in terms of trade (trade a certain category of goods in a volume that exceeds the 90^{th} percentile of the sample); 2- municipalities that traded a certain category of products at least once in the period from 2010-2017; 3- total sample of municipalities but excluding those that border other countries or bordering the water-front directly. And finally, I re-estimate the results for the universe of immigrants as per section 5.2 but including baseline controls (2000 or 2010 depending on availability of data and interacted with a time dummy).

Table II.46: **Excluding Outlier Municipalities:** Total Exports and Imports

	(1) log(Total Exports)	(2) log(Total Exports)	(3) log(Total Imports)	(4) log(Total Imports)
Migrants	-0.0140	0.0274	0.00450	0.274
	(0.0122)	(0.290)	(0.00782)	(0.238)
Observations	42315	42293	42315	42300
Estimation	OLS	IV	OLS	IV
Kleibergen-Paap F-stat		18.8440		17.5518
Mean Migrants	0.000576	0.000576	0.000557	0.000557
Doubling of Mean	-0.000807	0.00158	0.000251	0.0153
Municipality Controls	Yes	Yes	Yes	Yes
Year F.E	Yes	Yes	Yes	Yes
Municipality F.E	Yes	Yes	Yes	Yes

Notes: The table shows OLS and 2SLS results for total export and imports, excluding outlier municipalities in terms of the volume of trade. The outlier municipalities are those that trade more than the 90 percentile of the sample. The dependent variable is of the form $log(x+1)$ and the explanatory variable of immigration is the ratio of immigrants to the population per municipality (in %). Robust standard errors in parentheses clustered at the mesoregion level. The predicted migration variable is divided by population in each municipality. Data sources: IBGE data for control variables, COMEX for trade data and SINCRE for migration data. Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1.

Table II.47: **Excluding Outlier Municipalities:** Differentiated, Homogeneous, and Reference Exports based on Rauch Classification

	(1) log(Differentiated Exports)	(2) log(Differentiated Exports)	(3) log(Homogeneous Exports)	(4) log(Homogeneous Exports)
Migrants	-0.0135	-0.101	-0.0139	0.115
	(0.0122)	(0.267)	(0.0128)	(0.228)
Observations	42315	42294	42315	42285
Estimation	OLS	IV	OLS	IV
Kleibergen-Paap F-stat		18.2392		19.4617
Mean Migrants	0.000573	0.000573	0.000584	0.000584
Doubling of Mean	-0.000776	-0.00581	-0.000812	0.00671
Municipality Controls	Yes	Yes	Yes	Yes
Year F.E	Yes	Yes	Yes	Yes
Municipality F.E	Yes	Yes	Yes	Yes

Notes: The table shows OLS and 2SLS results for subcategories of exports based on Rauch Classification of goods, excluding outlier municipalities in terms of the volume of trade. The outlier municipalities are those that trade more than the 90 percentile of the sample. The dependent variable is of the form $log(x + 1)$ and the explanatory variable of immigration is the ratio of immigrants to the population per municipality (in %). Robust standard errors in parentheses clustered at the mesoregion level. The predicted migration variable is divided by population in each municipality. Data sources: IBGE data for control variables, COMEX for trade data and SINCRE for migration data. Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1.

Table II.48: **Excluding Outlier Municipalities:** Differentiated, Homogeneous, and Reference Imports based on Rauch Classification

	(1) log(Differentiated Imports)	(2) log(Differentiated Imports)	(3) log(Homogeneous Imports)	(4) log(Homogeneous Imports)
Migrants	-0.000363	0.249	-0.00841	0.0300
	(0.00442)	(0.204)	(0.0129)	(0.219)
Observations	42315	42291	42315	42293
Estimation	OLS	IV	OLS	IV
Kleibergen-Paap F-stat		17.8140		16.5921
Mean Migrants	0.000555	0.000555	0.000563	0.000563
Doubling of Mean	-0.0000201	0.0138	-0.000473	0.00169
Municipality Controls	Yes	Yes	Yes	Yes
Year F.E	Yes	Yes	Yes	Yes
Municipality F.E	Yes	Yes	Yes	Yes

Notes: The table shows OLS and 2SLS results for subcategories of imports based on Rauch Classification of goods, excluding outlier municipalities in terms of the volume of trade. The outlier municipalities are those that trade more than the 90 percentile of the sample. The dependent variable is of the form $log(x + 1)$ and the explanatory variable of immigration is the ratio of immigrants to the population per municipality (in %). Robust standard errors in parentheses clustered at the mesoregion level. The predicted migration variable is divided by population in each municipality. Data sources: IBGE data for control variables, COMEX for trade data and SINCRE for migration data. Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1.

Table II.49: **Excluding Outlier Municipalities:** Based on σ Elasticity of Substitution

	(1) log(Exports Low Diff.)	(2) log(Exports Low Diff.)	(3) log(Exports Mid Diff.)	(4) log(Exports Mid Diff.)	(5) log(Exports High Diff.)	(6) log(Exports High Diff.)
Migrants	-0.0106	-0.0434	-0.0185	0.0336	0.00360	0.221
	(0.0124)	(0.297)	(0.0163)	(0.272)	(0.00701)	(0.264)
Observations	42315	42291	42315	42285	42315	42289
Estimation	OLS	IV	OLS	IV	OLS	IV
Kleibergen-Paap F-stat		17.3586		18.4494		18.4408
Mean Migrants	0.000567	0.000567	0.000591	0.000591	0.000574	0.000574
Doubling of Mean	-0.000603	-0.00247	-0.00109	0.00198	0.000207	0.0127
	(1) log(Import Low Diff.)	(2) log(Import Low Diff.)	(3) log(Import Mid Diff.)	(4) log(Import Mid Diff.)	(5) log(Import High Diff.)	(6) log(Import High Diff.)
Migrants	0.00516	0.261	0.00202	0.165	-0.0124	0.248
	(0.00558)	(0.224)	(0.00645)	(0.200)	(0.0139)	(0.218)
Observations	42315	42300	42315	42295	42315	42293
Estimation	OLS	IV	OLS	IV	OLS	IV
Kleibergen-Paap F-stat		17.2035		17.7778		16.0307
Mean Migrants	0.000551	0.000551	0.000559	0.000559	0.000553	0.000553
Doubling of Mean	0.000285	0.0144	0.000113	0.00919	-0.000687	0.0137
Municipality Controls	Yes	Yes	Yes	Yes	Yes	Yes
Year F.E	Yes	Yes	Yes	Yes	Yes	Yes
Municipality F.E	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The table shows OLS and 2SLS results for subcategories of exports and imports based on σ elasticity of substitution, excluding outlier municipalities in terms of the volume of trade. The outlier municipalities are those that trade more than the 90 percentile of the sample. The dependent variable is of the form $log(x+1)$ and the explanatory variable of immigration is the ratio of immigrants to the population per municipality (in %). Robust standard errors in parentheses clustered at the mesoregion level. The predicted migration variable is divided by population in each municipality. Data sources: IBGE data for control variables, COMEX for trade data and SINCRE for migration data. Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1.

Table II.50: **Excluding non-Trading Municipalities:** Total Exports and Imports

	(1) log(Total Exports)	(2) log(Total Exports)	(3) log(Total Imports)	(4) log(Total Imports)
Migrants	-0.0824**	-0.194	0.0762*	0.753
	(0.0355)	(0.512)	(0.0446)	(0.471)
Observations	15364	15083	14833	14470
Estimation	OLS	IV	OLS	IV
Kleibergen-Paap F-stat		12.1023		14.1069
Mean Migrants	0.0973	0.0973	0.107	0.107
Doubling of Mean	-0.00801	-0.0189	0.00818	0.0809
Municipality Controls	Yes	Yes	Yes	Yes
Year F.E	Yes	Yes	Yes	Yes
Municipality F.E	Yes	Yes	Yes	Yes

Notes: The table shows OLS and 2SLS results for total export and imports, excluding municipalities that do not trade. The dependent variable is of the form $log(x+1)$ and the explanatory variable of immigration is the ratio of immigrants to the population per municipality (in %). Robust standard errors in parentheses clustered at the mesoregion level. The predicted migration variable is divided by population in each municipality. Data sources: IBGE data for control variables, COMEX for trade data and SINCRE for migration data. Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1.

Table II.51: **Excluding non-Trading Municipalities:** Differentiated, Homogeneous, and Reference Exports based on Rauch Classification

	(1) log(Differentiated Exports)	(2) log(Differentiated Exports)	(3) log(Homogeneous Exports)	(4) log(Homogeneous Exports)
Migrants	-0.0779*	-0.102	-0.0515	-0.0316
	(0.0415)	(0.510)	(0.0411)	(0.560)
Observations	11479	11153	10530	10260
Estimation	OLS	IV	OLS	IV
Kleibergen-Paap F-stat		9.2833		8.2770
Mean Migrants	0.114	0.114	0.116	0.116
Doubling of Mean	-0.00887	-0.0117	-0.00596	-0.00365
Municipality Controls	Yes	Yes	Yes	Yes
Year F.E	Yes	Yes	Yes	Yes
Municipality F.E	Yes	Yes	Yes	Yes

Notes: The table shows OLS and 2SLS results for subcategories of exports based on Rauch Classification of goods, excluding municipalities that do not trade. The dependent variable is of the form $log(x + 1)$ and the explanatory variable of immigration is the ratio of immigrants to the population per municipality (in %). Robust standard errors in parentheses clustered at the mesoregion level. The predicted migration variable is divided by population in each municipality. Data sources: IBGE data for control variables, COMEX for trade data and SINCRE for migration data. Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1.

Table II.52: **Excluding non-Trading Municipalities:** Differentiated, Homogeneous, and Reference Imports based on Rauch Classification

	(1) log(Differentiated Imports)	(2) log(Differentiated Imports)	(3) log(Homogeneous Imports)	(4) log(Homogeneous Imports)
Migrants	0.0670**	0.709	0.0194	-0.621
	(0.0304)	(0.503)	(0.0566)	(0.653)
Observations	13401	13006	10725	10465
Estimation	OLS	IV	OLS	IV
Kleibergen-Paap F-stat		12.6036		11.3140
Mean Migrants	0.113	0.113	0.125	0.125
Doubling of Mean	0.00755	0.0799	0.00242	-0.0773
Municipality Controls	Yes	Yes	Yes	Yes
Year F.E	Yes	Yes	Yes	Yes
Municipality F.E	Yes	Yes	Yes	Yes

Notes: The table shows OLS and 2SLS results for subcategories of imports based on Rauch Classification of goods, excluding municipalities that do not trade. The dependent variable is of the form $log(x + 1)$ and the explanatory variable of immigration is the ratio of immigrants to the population per municipality (in %). Robust standard errors in parentheses clustered at the mesoregion level. The predicted migration variable is divided by population in each municipality. Data sources: IBGE data for control variables, COMEX for trade data and SINCRE for migration data. Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1.

Table II.53: **Excluding non-Trading Municipalities:** Based on σ Elasticity of Substitution

	(1) log(Exports Low Diff.)	(2) log(Exports Low Diff.)	(3) log(Exports Mid Diff.)	(4) log(Exports Mid Diff.)	(5) log(Exports High Diff.)	(6) log(Exports High Diff.)
Migrants	-0.0553	-0.287	-0.116**	0.0779	0.00817	-0.160
	(0.0692)	(0.607)	(0.0503)	(0.552)	(0.0655)	(0.620)
Observations	11694	11399	11357	11043	7413	7150
Estimation	OLS	IV	OLS	IV	OLS	IV
Kleibergen-Paap F-stat		8.8754		9.1002		7.5877
Mean Migrants	0.116	0.116	0.115	0.115	0.137	0.137
Doubling of Mean	-0.00640	-0.0331	-0.0133	0.00894	0.00112	-0.0219
	(1) log(Import Low Diff.)	(2) log(Import Low Diff.)	(3) log(Import Mid Diff.)	(4) log(Import Mid Diff.)	(5) log(Import High Diff.)	(6) log(Import High Diff.)
Migrants	0.102***	0.0114	0.0576	0.527	-0.103**	0.540
	(0.0388)	(0.513)	(0.0456)	(0.495)	(0.0405)	(0.631)
Observations	13009	12639	12685	12374	8819	8526
Estimation	OLS	IV	OLS	IV	OLS	IV
Kleibergen-Paap F-stat		11.0481		12.2174		7.6769
Mean Migrants	0.113	0.113	0.117	0.117	0.138	0.138
Doubling of Mean	0.0115	0.00129	0.00675	0.0618	-0.0143	0.0748
Municipality Controls	Yes	Yes	Yes	Yes	Yes	Yes
Year F.E	Yes	Yes	Yes	Yes	Yes	Yes
Municipality F.E	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The table shows OLS and 2SLS results for subcategories of exports and imports based on σ elasticity of substitution, excluding municipalities that do not trade. The dependent variable is of the form $log(x+1)$ and the explanatory variable of immigration is the ratio of immigrants to the population per municipality (in %). Robust standard errors in parentheses clustered at the mesoregion level. The predicted migration variable is divided by population in each municipality. Data sources: IBGE data for control variables, COMEX for trade data and SINCRE for migration data. Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1.

Table II.54: **Excluding Border Municipalities:** Total Exports and Imports

	(1) log(Total Exports)	(2) log(Total Exports)	(3) log(Total Imports)	(4) log(Total Imports)
Migrants	-0.0123	-0.198	0.000847	0.585*
	(0.0117)	(0.374)	(0.00549)	(0.336)
Observations	39840	39840	39840	39840
Estimation	OLS	IV	OLS	IV
Kleibergen-Paap F-stat		15.2766		15.2873
Mean Migrants	0.000562	0.000562	0.000562	0.000562
Doubling of Mean	-0.000693	-0.0111	0.0000476	0.0329
Municipality Controls	Yes	Yes	Yes	Yes
Year F.E	Yes	Yes	Yes	Yes
Municipality F.E	Yes	Yes	Yes	Yes

Notes: The table shows OLS and 2SLS results for total export and imports, excluding municipalities that lie on the borders of Brazil. Location data on municipalities at the border are retrieved from IBGE. The dependent variable is of the form $log(x+1)$ and the explanatory variable of immigration is the ratio of immigrants to the population per municipality (in %). Robust standard errors in parentheses clustered at the mesoregion level. The predicted migration variable is divided by population in each municipality. Data sources: IBGE data for control variables, COMEX for trade data and SINCRE for migration data. Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1.

Table II.55: **Excluding Border Municipalities:** Differentiated, Homogeneous, and Reference Exports based on Rauch Classification

	(1) log(Differentiated Exports)	(2) log(Differentiated Exports)	(3) log(Homogeneous Exports)	(4) log(Homogeneous Exports)
Migrants	-0.00667	-0.296	-0.0125	-0.0764
	(0.00885)	(0.372)	(0.0127)	(0.320)
Observations	39840	39840	39840	39840
Estimation	OLS	IV	OLS	IV
Kleibergen-Paap F-stat		15.2859		15.2777
Mean Migrants	0.000562	0.000562	0.000562	0.000562
Doubling of Mean	-0.000375	-0.0166	-0.000705	-0.00429
Municipality Controls	Yes	Yes	Yes	Yes
Year F.E	Yes	Yes	Yes	Yes
Municipality F.E	Yes	Yes	Yes	Yes

Notes: The table shows OLS and 2SLS results for subcategories of exports based on Rauch Classification of goods, excluding municipalities that lie on the borders of Brazil. Location data on municipalities at the border are retrieved from IBGE. The dependent variable is of the form $log(x + 1)$ and the explanatory variable of immigration is the ratio of immigrants to the population per municipality (in %). Robust standard errors in parentheses clustered at the mesoregion level. The predicted migration variable is divided by population in each municipality. Data sources: IBGE data for control variables, COMEX for trade data and SINCRE for migration data. Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1.

Table II.56: **Excluding Border Municipalities:** Differentiated, Homogeneous, and Reference Imports based on Rauch Classification

	(1) log(Differentiated Imports)	(2) log(Differentiated Imports)	(3) log(Homogeneous Imports)	(4) log(Homogeneous Imports)
Migrants	0.00241	0.679**	-0.00880	0.227
	(0.00391)	(0.337)	(0.0131)	(0.267)
Observations	39840	39840	39840	39840
Estimation	OLS	IV	OLS	IV
Kleibergen-Paap F-stat		15.2889		15.2746
Mean Migrants	0.000562	0.000562	0.000562	0.000562
Doubling of Mean	0.000135	0.0382	-0.000495	0.0128
Municipality Controls	Yes	Yes	Yes	Yes
Year F.E	Yes	Yes	Yes	Yes
Municipality F.E	Yes	Yes	Yes	Yes

Notes: The table shows OLS and 2SLS results for subcategories of imports based on Rauch Classification of goods, excluding municipalities that lie on the borders of Brazil. Location data on municipalities at the border are retrieved from IBGE. The dependent variable is of the form $log(x + 1)$ and the explanatory variable of immigration is the ratio of immigrants to the population per municipality (in %). Robust standard errors in parentheses clustered at the mesoregion level. The predicted migration variable is divided by population in each municipality. Data sources: IBGE data for control variables, COMEX for trade data and SINCRE for migration data. Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1.

Table II.57: **Excluding Border Municipalities:** Based on σ Elasticity of Substitution

	(1) log(Exports Low Diff.)	(2) log(Exports Low Diff.)	(3) log(Exports Mid Diff.)	(4) log(Exports Mid Diff.)	(5) log(Exports High Diff.)	(6) log(Exports High Diff.)
Migrants	-0.00191	-0.0185	-0.0149	-0.186	0.00484	0.0796
	(0.00711)	(0.379)	(0.0147)	(0.385)	(0.00706)	(0.336)
Observations	39840	39840	39840	39840	39840	39840
Estimation	OLS	IV	OLS	IV	OLS	IV
Kleibergen-Paap F-stat		15.2779		15.2267		15.2837
Mean Migrants	0.000562	0.000562	0.000562	0.000562	0.000562	0.000562
Doubling of Mean	-0.000108	-0.00104	-0.000840	-0.0105	0.000272	0.00448
	(1) log(Import Low Diff.)	(2) log(Import Low Diff.)	(3) log(Import Mid Diff.)	(4) log(Import Mid Diff.)	(5) log(Import High Diff.)	(6) log(Import High Diff.)
Migrants	0.00383	0.660**	-0.00184	0.574*	0.00263	0.559
	(0.00433)	(0.324)	(0.00574)	(0.315)	(0.00569)	(0.341)
Observations	39840	39840	39840	39840	39840	39840
Estimation	OLS	IV	OLS	IV	OLS	IV
Kleibergen-Paap F-stat		15.2847		15.2891		15.2906
Mean Migrants	0.000562	0.000562	0.000562	0.000562	0.000562	0.000562
Doubling of Mean	0.000215	0.0371	-0.000103	0.0323	0.000148	0.0315
Municipality Controls	Yes	Yes	Yes	Yes	Yes	Yes
Year F.E	Yes	Yes	Yes	Yes	Yes	Yes
Municipality F.E	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The table shows OLS and 2SLS results for subcategories of exports and imports based on σ elasticity of substitution, excluding municipalities that lie on the borders of Brazil. Location data on municipalities at the border are retrieved from IBGE. The dependent variable is of the form $log(x+1)$ and the explanatory variable of immigration is the ratio of immigrants to the population per municipality (in %). Robust standard errors in parentheses clustered at the mesoregion level. The predicted migration variable is divided by population in each municipality. Data sources: IBGE data for control variables, COMEX for trade data and SINCRE for migration data. Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1.

Table II.58: **Controls in year 2000 interacted with a time trend:** Total Exports and Imports

	(1) log(Total Exports)	(2) log(Total Exports)	(3) log(Total Imports)	(4) log(Total Imports)
Migrants	-0.0148	0.126	0.00639	0.363
	(0.0128)	(0.276)	(0.00946)	(0.231)
Observations	44545	44545	44545	44545
Estimation	OLS	IV	OLS	IV
Kleibergen-Paap F-stat		23.5423		23.5752
Mean Migrants	0.000633	0.000633	0.000633	0.000633
Doubling of Mean	-0.000934	0.00795	0.000404	0.0230
Municipality Controls	Yes	Yes	Yes	Yes
Year F.E	Yes	Yes	Yes	Yes
Municipality F.E	Yes	Yes	Yes	Yes

Notes: The table shows OLS and 2SLS results for total export and imports. This specification includes controls in 2010 interacted with a time trend. The dependent variable is of the form $log(x+1)$ and the explanatory variable of immigration is the ratio of immigrants to the population per municipality (in %). Robust standard errors in parentheses clustered at the mesoregion level. The predicted migration variable is divided by population in each municipality. Data sources: IBGE data for control variables, COMEX for trade data and SINCRE for migration data. Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1.

Table II.59: **Controls in year 2000 interacted with a time trend:** Differentiated, Homogeneous, and Reference Exports based on Rauch Classification

	(1) log(Differentiated Exports)	(2) log(Differentiated Exports)	(3) log(Homogeneous Exports)	(4) log(Homogeneous Exports)
Migrants	-0.00997	-0.0931	-0.0146	0.252
	(0.0108)	(0.259)	(0.0131)	(0.229)
Observations	44545	44545	44545	44545
Estimation	OLS	IV	OLS	IV
Kleibergen-Paap F-stat		23.5552		23.5267
Mean Migrants	0.000633	0.000633	0.000633	0.000633
Doubling of Mean	-0.000631	-0.00589	-0.000923	0.0160
Municipality Controls	Yes	Yes	Yes	Yes
Year F.E	Yes	Yes	Yes	Yes
Municipality F.E	Yes	Yes	Yes	Yes

Notes: The table shows OLS and 2SLS results for subcategories of exports based on Rauch Classification of goods. This specification includes controls in 2010 interacted with a time trend. The dependent variable is of the form $log(x+1)$ and the explanatory variable of immigration is the ratio of immigrants to the population per municipality (in %). Robust standard errors in parentheses clustered at the mesoregion level. The predicted migration variable is divided by population in each municipality. Data sources: IBGE data for control variables, COMEX for trade data and SINCRE for migration data. Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1.

Table II.60: **Controls in year 2000 interacted with a time trend:** Differentiated, Homogeneous, and Reference Imports based on Rauch Classification

	(1) log(Differentiated Imports)	(2) log(Differentiated Imports)	(3) log(Homogeneous Imports)	(4) log(Homogeneous Imports)
Migrants	0.00249	0.458**	-0.00922	0.0530
	(0.00537)	(0.231)	(0.0143)	(0.205)
Observations	44545	44545	44545	44545
Estimation	OLS	IV	OLS	IV
Kleibergen-Paap F-stat		23.5628		23.5534
Mean Migrants	0.000633	0.000633	0.000633	0.000633
Doubling of Mean	0.000157	0.0290	-0.000584	0.00335
Municipality Controls	Yes	Yes	Yes	Yes
Year F.E	Yes	Yes	Yes	Yes
Municipality F.E	Yes	Yes	Yes	Yes

Notes: The table shows OLS and 2SLS results for subcategories of imports based on Rauch Classification of goods. This specification includes controls in 2010 interacted with a time trend. The dependent variable is of the form $log(x+1)$ and the explanatory variable of immigration is the ratio of immigrants to the population per municipality (in %). Robust standard errors in parentheses clustered at the mesoregion level. The predicted migration variable is divided by population in each municipality. Data sources: IBGE data for control variables, COMEX for trade data and SINCRE for migration data. Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1.

Table II.61: **Controls in year 2000 interacted with a time trend:** Based on σ Elasticity of Substitution

	(1) log(Exports Low Diff.)	(2) log(Exports Low Diff.)	(3) log(Exports Mid Diff.)	(4) log(Exports Mid Diff.)	(5) log(Exports High Diff.)	(6) log(Exports High Diff.)
Migrants	-0.00956	0.151	-0.0187	0.0175	0.00881	0.249
	(0.0122)	(0.285)	(0.0165)	(0.282)	(0.0103)	(0.249)
Observations	44545	44545	44545	44545	44545	44545
Estimation	OLS	IV	OLS	IV	OLS	IV
Kleibergen-Paap F-stat		23.5247		23.4402		23.5590
Mean Migrants	0.000633	0.000633	0.000633	0.000633	0.000633	0.000633
Doubling of Mean	-0.000605	0.00957	-0.00118	0.00111	0.000557	0.0158
	(1) log(Import Low Diff.)	(2) log(Import Low Diff.)	(3) log(Import Mid Diff.)	(4) log(Import Mid Diff.)	(5) log(Import High Diff.)	(6) log(Import High Diff.)
Migrants	0.0120	0.360	0.00371	0.354	-0.0145	0.224
	(0.0118)	(0.234)	(0.00834)	(0.216)	(0.0138)	(0.226)
Observations	44545	44545	44545	44545	44545	44545
Estimation	OLS	IV	OLS	IV	OLS	IV
Kleibergen-Paap F-stat		23.5655		23.5706		23.5355
Mean Migrants	0.000633	0.000633	0.000633	0.000633	0.000633	0.000633
Doubling of Mean	0.000757	0.0228	0.000235	0.0224	-0.000915	0.0142
Municipality Controls	Yes	Yes	Yes	Yes	Yes	Yes
Year F.E	Yes	Yes	Yes	Yes	Yes	Yes
Municipality F.E	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The table shows OLS and 2SLS results for subcategories of exports and imports based on σ elasticity of substitution. This specification includes controls in 2010 interacted with a time trend. The dependent variable is of the form $log(x+1)$ and the explanatory variable of immigration is the ratio of immigrants to the population per municipality (in %). Robust standard errors in parentheses clustered at the mesoregion level. The predicted migration variable is divided by population in each municipality. Data sources: IBGE data for control variables, COMEX for trade data and SINCRE for migration data. Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1.

II.C.6 Additional Results of 2SLS Estimations for Different Country Samples

In what follows, I present the results estimated using equation (5.15) but varying the sample of countries in the analysis. I present three different samples: For Mercosur countries, OECD countries and countries that had 15 or no immigration presence prior to 2000, but had higher immigration flows (more than a 100) between 2010-2017. The latter set of countries is composed of 57 countries and includes Azerbaijan, Bahamas, Bahrain, Barbados, Belize, Benin, Bermuda, Bhutan, Bosnia and Herzegovina, Botswana, Brunei, Burkina Faso, Burundi, Cambodia, Central African Republic, Chad, Comoros, Republic of Congo, Cyprus, Djibouti, East Timor, Equatorial Guinea, Eritrea, Estonia, Ethiopia, Fiji, French Polynesia, Gabon, Gambia, Georgia, Gibraltar, Grenada, Guadeloupe, Guyana, Iceland, Jamaica, Kazakhstan, Kyrgyz Republic, Lesotho, Liechtenstein , Luxembourg, Macedonia, Madagascar, Malawi, Maldives, Mali, Malta, Marshall Islands, Mauritania, Mauritius, Monaco, Mongolia, Myanmar, Namibia, Nepal, New Caledonia, Niger, Oman, Papua New Guinea, Qatar, Samoa, San Marino, Sao Tome and Principe, Saudi Arabia, Seychelles, Somalia, Sri Lanka, St.

Lucia, Swaziland, Tajikistan, Togo, Turkmenistan, Uganda, Uzbekistan, Yemen, and Zambia.

Table II.62: **Less than 15 immigrants historically:** Total Exports and Imports

	(1) log(Total Exports in USD)	(2) log(Total Exports in USD)	(3) log(Total Imports in USD)	(4) log(Total Imports in USD)
Migrants	-0.357	81.32*	-0.121	31.99
	(0.286)	(42.84)	(0.212)	(29.65)
Observations	44543	44543	44543	44543
Estimation	OLS	IV	OLS	IV
Mean Migrants	0.00000556	0.00000556	0.00000556	0.00000556
Doubling of Mean	-0.000198	0.0452	-0.0000672	0.0178
Municipality Controls	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Municipality F.E	Yes	Yes	Yes	Yes

Notes: The table shows the OLS and 2SLS estimates for the impact of immigration on exports and imports using only the sample of countries that historically (prior to 2000) only had less than 15 immigrants. The dependent variable is of the form $log(x+1)$ and the explanatory variable of immigration is the ratio of immigrants to the population per municipality (in %). Robust standard errors in parentheses clustered at the mesoregion level. Regressions include mesoregion fixed effects and municipality controls. The predicted migration variable is divided by population in each municipality. Data sources: IBGE data for control variables, COMEX for trade data and SINCRE for migration data. Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1.

Table II.63: **Less than 15 immigrants historically:** Differentiated and Homogeneous Exports based on Rauch Classification

	(1) log(Differentiated Exports in USD)	(2) log(Differentiated Exports in USD)	(3) log(Homogeneous Exports in USD)	(4) log(Homogeneous Exports in USD)
Migrants	0.196	22.90	-0.264	95.33**
	(0.214)	(28.31)	(0.258)	(47.23)
Observations	44543	44543	44543	44543
Estimation	OLS	IV	OLS	IV
Mean Migrants	0.00000556	0.00000556	0.00000556	0.00000556
Doubling of Mean	0.000109	0.0127	-0.000147	0.0530
Municipality Controls	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Municipality F.E	Yes	Yes	Yes	Yes

Notes: The table shows the OLS and 2SLS estimates for the impact of immigration on differentiated and homogeneous exports using only the sample of countries that historically (prior to 2000) only had less than 15 immigrants. The dependent variable is of the form $log(x + 1)$ and the explanatory variable of immigration is the ratio of immigrants to the population per municipality (in %). Robust standard errors in parentheses clustered at the mesoregion level. Regressions include mesoregion fixed effects and municipality controls. The predicted migration variable is divided by population in each municipality. Data sources: IBGE data for control variables, COMEX for trade data and SINCRE for migration data. Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1.

Table II.64: **Less than 15 immigrants historically:** Differentiated and Homogeneous Imports based on Rauch Classification

	(1) log(Differentiated Imports in USD)	(2) log(Differentiated Imports in USD)	(3) log(Homogeneous Imports in USD)	(4) log(Homogeneous Imports in USD)
Migrants	-0.234	-23.42	-0.240	17.23
	(0.219)	(27.81)	(0.214)	(26.60)
Observations	44543	44543	44543	44543
Estimation	OLS	IV	OLS	IV
Mean Migrants	0.00000556	0.00000556	0.00000556	0.00000556
Doubling of Mean	-0.000130	-0.0130	-0.000133	0.00957
Municipality Controls	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Municipality F.E	Yes	Yes	Yes	Yes

Notes: The table shows the OLS and 2SLS estimates for the impact of immigration on differentiated and homogeneous imports using only the sample of countries that historically (prior to 2000) only had less than 15 immigrants. The dependent variable is of the form $log(x + 1)$ and the explanatory variable of immigration is the ratio of immigrants to the population per municipality (in %). Robust standard errors in parentheses clustered at the mesoregion level. Regressions include mesoregion fixed effects and municipality controls. The predicted migration variable is divided by population in each municipality. Data sources: IBGE data for control variables, COMEX for trade data and SINCRE for migration data. Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1.

Table II.65: **Less than 15 immigrants historically:** Differentiated and Homogeneous Exports and Imports based on Rauch Classification on the Extensive Margin

	(1) Total Exports in USD	(2) Differentiated Exports in USD	(3) Homogeneous Exports in USD	(4) Total Imports in USD	(5) Differentiated Imports in USD	(6) Homogeneous Imports in USD
Migrants	-24.74	21.27	62.17	3.961	-4.493	-1.986
	(41.12)	(30.75)	(40.83)	(33.68)	(28.61)	(24.83)
Observations	44543	44543	44543	44543	44543	44543
Estimation	IV	IV	IV	IV	IV	IV
Doubling of Mean	-0.0137	0.0118	0.0345	0.00220	-0.00250	-0.00110
Municipality Controls	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Municipality F.E	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The table shows the OLS and 2SLS estimates for the impact of immigration, at the extensive margin, on differentiated and homogeneous exports and imports using only the sample of countries that historically (prior to 2000) only had less than 15 immigrants. The dependent variable is of the form $log(x+1)$ and the explanatory variable of immigration is the ratio of immigrants to the population per municipality (in %). Robust standard errors in parentheses clustered at the mesoregion level. Regressions include mesoregion fixed effects and municipality controls. The predicted migration variable is divided by population in each municipality. Data sources: IBGE data for control variables, COMEX for trade data and SINCRE for migration data. Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1.

Table II.66: **Less than 15 immigrants historically:** Based on σ Elasticity of Substitution

	(1) log(Exports Low Diff.)	(2) log(Exports Low Diff.)	(3) log(Exports Mid Diff.)	(4) log(Exports Mid Diff.)	(5) log(Exports High Diff.)	(6) log(Exports High Diff.)
Migrants	-0.408	43.35	-0.0298	44.53	0.306	60.94*
	(0.289)	(42.67)	(0.162)	(33.59)	(0.250)	(34.93)
Observations	44543	44543	44543	44543	44543	44543
Estimation	OLS	IV	OLS	IV	OLS	IV
Mean Migrants	0.00000556	0.00000556	0.00000556	0.00000556	0.00000556	0.00000556
Doubling of Mean	-0.000227	0.0241	-0.0000166	0.0247	0.000170	0.0339
	(1) log(Import Low Diff.)	(2) log(Import Low Diff.)	(3) log(Import Mid Diff.)	(4) log(Import Mid Diff.)	(5) log(Import High Diff.)	(6) log(Import High Diff.)
Migrants	-0.179	-41.90	0.0000588	87.82**	-0.142	13.72
	(0.193)	(28.45)	(0.212)	(38.64)	(0.223)	(16.61)
Observations	44543	44543	44543	44543	44543	44543
Estimation	OLS	IV	OLS	IV	OLS	IV
Mean Migrants	0.00000556	0.00000556	0.00000556	0.00000556	0.00000556	0.00000556
Doubling of Mean	-0.0000996	-0.0233	3.27e-08	0.0488	-0.0000790	0.00762
Municipality Controls	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Municipality F.E	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The table shows 2SLS results for subcategories of exports and imports based on the elasticity of substitution σ using only the sample of countries that historically (prior to 2000) only had less than 15 immigrants. Higher elasticity of substitution means that the product is less differentiated and hence more homogeneous. The dependent variable is of the form $log(x+1)$ and the explanatory variable of immigration is the ratio of immigrants to the population per municipality (in %). Robust standard errors in parentheses clustered at the mesoregion level. Regressions include mesoregion fixed effects and municipality controls. The predicted migration variable is divided by population in each municipality. Data sources: IBGE data for control variables, COMEX for trade data and SINCRE for migration data. Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1.

Table II.67: **Less than 15 immigrants historically:** Based on σ Elasticity of Substitution on the Extensive margin

	(1) Exports Low Diff.	(2) Exports Mid Diff.	(3) Exports High Diff.	(4) Import Low Diff.	(5) Import Mid Diff.	(6) Import High Diff.
Migrants	23.04	-4.069	18.72	-17.09	32.58	11.08
	(42.14)	(36.07)	(28.75)	(31.23)	(35.40)	(15.71)
Observations	44543	44543	44543	44543	44543	44543
Estimation	IV	IV	IV	IV	IV	IV
Doubling of Mean	0.0128	-0.00226	0.0104	-0.00950	0.0181	0.00615
Municipality Controls	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Municipality F.E	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The table shows 2SLS results for subcategories of exports and imports, at the extensive margin, based on the elasticity of substitution σ using only the sample of countries that historically (prior to 2000) only had less than 15 immigrants. Higher elasticity of substitution means that the product is less differentiated and hence more homogeneous. The dependent variable is of the form $log(x+1)$ and the explanatory variable of immigration is the ratio of immigrants to the population per municipality (in %). Robust standard errors in parentheses clustered at the mesoregion level. Regressions include mesoregion fixed effects and municipality controls in 2010. The predicted migration variable is divided by population in each municipality. Data sources: IBGE data for control variables, COMEX for trade data and SINCRE for migration data. Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1.

Table II.68: **Mercosur:** Total Exports and Imports

	(1) log(Total Exports)	(2) log(Total Exports)	(3) log(Total Imports)	(4) log(Total Imports)
Mercosur Migrants	0.00729	0.296	-0.0162	-0.413
	(0.0391)	(0.317)	(0.0131)	(0.298)
Observations	44543	44543	44543	44543
Estimation	OLS	IV	OLS	IV
Mean Migrants	0.000239	0.000239	0.000239	0.000239
Doubling of Mean	0.000174	0.00709	-0.000387	-0.00989
Municipality Controls	Yes	Yes	Yes	Yes
Year F.E	Yes	Yes	Yes	Yes
Municipality F.E	Yes	Yes	Yes	Yes

Notes: The table shows the OLS and 2SLS estimates for the impact of immigration on exports and imports using the sample of Mercosur countries. The dependent variable is of the form $log(x+1)$ and the explanatory variable of immigration is the ratio of immigrants to the population per municipality (in %). Robust standard errors in parentheses clustered at the mesoregion level. Regressions include mesoregion fixed effects and municipality controls. The predicted migration variable is divided by population in each municipality. Data sources: IBGE data for control variables, COMEX for trade data and SINCRE for migration data. Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1.

Table II.69: **Mercosur:** Differentiated and Homogeneous Exports and Imports based on Rauch Classification

	(1) log(Differentiated Exports)	(2) log(Differentiated Exports)	(3) log(Homogeneous Exports)	(4) log(Homogeneous Exports)
Mercosur Migrants	0.00161	0.0536	-0.0177	-0.475
	(0.0339)	(0.263)	(0.0183)	(0.318)
Observations	44543	44543	44543	44543
Estimation	OLS	IV	OLS	IV
Mean Migrants	0.000239	0.000239	0.000239	0.000239
Doubling of Mean	0.0000385	0.00128	-0.000425	-0.0114

	(1) log(Differentiated Imports)	(2) log(Differentiated Imports)	(3) log(Homogeneous Imports)	(4) log(Homogeneous Imports)
Mercosur Migrants	-0.0198	-0.481	-0.0246	-0.257
	(0.0150)	(0.321)	(0.0173)	(0.229)
Doubling of Mean	-0.000474	-0.0115	-0.000587	-0.00615
Municipality Controls	Yes	Yes	Yes	Yes
Year F.E	Yes	Yes	Yes	Yes
Municipality F.E	Yes	Yes	Yes	Yes

Notes: The table shows the OLS and 2SLS estimates for the impact of immigration on differentiated and homogeneous exports and imports using the sample of Mercosur countries. The dependent variable is of the form $log(x+1)$ and the explanatory variable of immigration is the ratio of immigrants to the population per municipality (in %). Robust standard errors in parentheses clustered at the mesoregion level. Regressions include mesoregion fixed effects and municipality controls. The predicted migration variable is divided by population in each municipality. Data sources: IBGE data for control variables, COMEX for trade data and SINCRE for migration data. Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1.

Table II.70: **Mercosur:** Based on σ Elasticity of Substitution

	(1) log(Exports Low Diff.)	(2) log(Exports Low Diff.)	(3) log(Exports Mid Diff.)	(4) log(Exports Mid Diff.)	(5) log(Exports High Diff.)	(6) log(Exports High Diff.)
Mercosur Migrants	-0.0270	0.572	-0.0214	-0.00212	0.0245	0.417
	(0.0317)	(0.374)	(0.0501)	(0.274)	(0.0216)	(0.277)
Observations	44543	44543	44543	44543	44543	44543
Estimation	OLS	IV	OLS	IV	OLS	IV
Mean Migrants	0.000239	0.000239	0.000239	0.000239	0.000239	0.000239
Doubling of Mean	-0.000645	0.0137	-0.000511	-0.0000507	0.000586	0.00998

	(1) log(Import Low Diff.)	(2) log(Import Low Diff.)	(3) log(Import Mid Diff.)	(4) log(Import Mid Diff.)	(5) log(Import High Diff.)	(6) log(Import High Diff.)
Mercosur Migrants	-0.00815	0.0371	-0.00623	-0.514	-0.0235	-0.449*
	(0.0129)	(0.275)	(0.0104)	(0.366)	(0.0171)	(0.249)
Doubling of Mean	-0.000195	0.000888	-0.000149	-0.0123	-0.000562	-0.0107
Municipality Controls	Yes	Yes	Yes	Yes	Yes	Yes
Year F.E	Yes	Yes	Yes	Yes	Yes	Yes
Municipality F.E	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The table shows 2SLS results for subcategories of exports and imports based on the elasticity of substitution σ using the sample of Mercosur countries. Higher elasticity of substitution means that the product is less differentiated and hence more homogeneous. The dependent variable is of the form $log(x+1)$ and the explanatory variable of immigration is the ratio of immigrants to the population per municipality (in %). Robust standard errors in parentheses clustered at the mesoregion level. Regressions include mesoregion fixed effects and municipality controls. The predicted migration variable is divided by population in each municipality. Data sources: IBGE data for control variables, COMEX for trade data and SINCRE for migration data. Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1.

Table II.71: **Mercosur:** Differentiated and Homogeneous Exports and Imports based on Rauch Classification at the Extensive Margin

	(1) Exports Low Diff.	(2) Exports Mid Diff.	(3) Exports High Diff.	(4) Import Low Diff.	(5) Import Mid Diff.	(6) Import High Diff.
Mercosur Migrants	0.562	0.318	0.293	-0.175	0.221	0.00404
	(0.360)	(0.350)	(0.223)	(0.267)	(0.338)	(0.198)
Observations	44543	44543	44543	44543	44543	44543
Estimation	IV	IV	IV	IV	IV	IV
Doubling of Mean	0.0134	0.00761	0.00701	-0.00418	0.00528	0.0000966
Municipality Controls	Yes	Yes	Yes	Yes	Yes	Yes
Year F.E	Yes	Yes	Yes	Yes	Yes	Yes
Municipality F.E	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The table shows the OLS and 2SLS estimates for the impact of immigration on differentiated and homogeneous exports and imports at the extensive margin using the sample of Mercosur countries. The dependent variable is of the form $log(x+1)$ and the explanatory variable of immigration is the ratio of immigrants to the population per municipality (in %). Robust standard errors in parentheses clustered at the mesoregion level. Regressions include mesoregion fixed effects and municipality controls. The predicted migration variable is divided by population in each municipality. Data sources: IBGE data for control variables, COMEX for trade data and SINCRE for migration data. Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1.

Table II.72: **Mercosur:** Based on σ Elasticity of Substitution at the Extensive Margin

	(1) log(Total Exports)	(2) log(Differentiated Exports)	(3) log(Homogeneous Exports)	(4) log(Total Imports)	(5) log(Differentiated Imports)	(6) log(Homogeneous Imports)
Mercosur Migrants	0.139	0.628	0.229	-0.166	-0.828**	0.193
	(0.295)	(0.388)	(0.322)	(0.351)	(0.395)	(0.250)
Observations	44543	44543	44543	44543	44543	44543
Estimation	IV	IV	IV	IV	IV	IV
Doubling of Mean	0.00334	0.0150	0.00549	-0.00398	-0.0198	0.00463
Municipality Controls	Yes	Yes	Yes	Yes	Yes	Yes
Year F.E	Yes	Yes	Yes	Yes	Yes	Yes
Municipality F.E	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The table shows 2SLS results for subcategories of exports and imports, at the extensive margin, based on the elasticity of substitution σ using the sample of Mercosur countries. Higher elasticity of substitution means that the product is less differentiated and hence more homogeneous. The dependent variable is of the form $log(x+1)$ and the explanatory variable of immigration is the ratio of immigrants to the population per municipality (in %). Robust standard errors in parentheses clustered at the mesoregion level. Regressions include mesoregion fixed effects and municipality controls. The predicted migration variable is divided by population in each municipality. Data sources: IBGE data for control variables, COMEX for trade data and SINCRE for migration data. Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1.

Table II.73: **OECD:** Differentiated and Homogeneous Exports and Imports based on Rauch Classification

	(1) log(Differentiated Exports)	(2) log(Differentiated Exports)	(3) log(Homogeneous Exports)	(4) log(Homogeneous Exports)
Mercosur Migrants	0.00161	0.0536	-0.0177	-0.475
	(0.0339)	(0.263)	(0.0183)	(0.318)
Observations	44543	44543	44543	44543
Estimation	OLS	IV	OLS	IV
Mean Migrants	0.000239	0.000239	0.000239	0.000239
Doubling of Mean	0.0000385	0.00128	-0.000425	-0.0114

	(1) log(Differentiated Imports)	(2) log(Differentiated Imports)	(3) log(Homogeneous Imports)	(4) log(Homogeneous Imports)
Mercosur Migrants	-0.0198	-0.481	-0.0246	-0.257
	(0.0150)	(0.321)	(0.0173)	(0.229)
Doubling of Mean	-0.000474	-0.0115	-0.000587	-0.00615
Municipality Controls	Yes	Yes	Yes	Yes
Year F.E	Yes	Yes	Yes	Yes
Municipality F.E	Yes	Yes	Yes	Yes

Notes: The table shows the OLS and 2SLS estimates for the impact of immigration on differentiated and homogeneous exports using the sample of OECD countries. The dependent variable is of the form $log(x+1)$ and the explanatory variable of immigration is the ratio of immigrants to the population per municipality (in %). Robust standard errors in parentheses clustered at the mesoregion level. Regressions include mesoregion fixed effects and municipality controls. The predicted migration variable is divided by population in each municipality. Data sources: IBGE data for control variables, COMEX for trade data and SINCRE for migration data. Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1.

Table II.74: **OECD:** Based on σ Elasticity of Substitution

	(1) log(Exports Low Diff.)	(2) log(Exports Low Diff.)	(3) log(Exports Mid Diff.)	(4) log(Exports Mid Diff.)	(5) log(Exports High Diff.)	(6) log(Exports High Diff.)
Mercosur Migrants	-0.0270	0.572	-0.0214	-0.00212	0.0245	0.417
	(0.0317)	(0.374)	(0.0501)	(0.274)	(0.0216)	(0.277)
Observations	44543	44543	44543	44543	44543	44543
Estimation	OLS	IV	OLS	IV	OLS	IV
Mean Migrants	0.000239	0.000239	0.000239	0.000239	0.000239	0.000239
Doubling of Mean	-0.000645	0.0137	-0.000511	-0.0000507	0.000586	0.00998

	(1) log(Import Low Diff.)	(2) log(Import Low Diff.)	(3) log(Import Mid Diff.)	(4) log(Import Mid Diff.)	(5) log(Import High Diff.)	(6) log(Import High Diff.)
Mercosur Migrants	-0.00815	0.0371	-0.00623	-0.514	-0.0235	-0.449*
	(0.0129)	(0.275)	(0.0104)	(0.366)	(0.0171)	(0.249)
Doubling of Mean	-0.000195	0.000888	-0.000149	-0.0123	-0.000562	-0.0107
Municipality Controls	Yes	Yes	Yes	Yes	Yes	Yes
Year F.E	Yes	Yes	Yes	Yes	Yes	Yes
Municipality F.E	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The table shows 2SLS results for subcategories of exports and imports based on the elasticity of substitution σ using the sample of OECD countries. Higher elasticity of substitution means that the product is less differentiated and hence more homogeneous. The dependent variable is of the form $log(x+1)$ and the explanatory variable of immigration is the ratio of immigrants to the population per municipality (in %). Robust standard errors in parentheses clustered at the mesoregion level. Regressions include mesoregion fixed effects and municipality controls. The predicted migration variable is divided by population in each municipality. Data sources: IBGE data for control variables, COMEX for trade data and SINCRE for migration data. Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1.

Table II.75: **OECD:** Differentiated and Homogeneous Exports and Imports based on Rauch Classification at the Extensive Margin

	(1) Exports Low Diff.	(2) Exports Mid Diff.	(3) Exports High Diff.	(4) Import Low Diff.	(5) Import Mid Diff.	(6) Import High Diff.
Mercosur Migrants	0.562	0.318	0.293	-0.175	0.221	0.00404
	(0.360)	(0.350)	(0.223)	(0.267)	(0.338)	(0.198)
Observations	44543	44543	44543	44543	44543	44543
Estimation	IV	IV	IV	IV	IV	IV
Doubling of Mean	0.0134	0.00761	0.00701	-0.00418	0.00528	0.0000966
Municipality Controls	Yes	Yes	Yes	Yes	Yes	Yes
Year F.E	Yes	Yes	Yes	Yes	Yes	Yes
Municipality F.E	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The table shows the OLS and 2SLS estimates for the impact of immigration on differentiated and homogeneous exports and imports at the extensive margin using the sample of OECD countries. The dependent variable is of the form $log(x+1)$ and the explanatory variable of immigration is the ratio of immigrants to the population per municipality (in %). Robust standard errors in parentheses clustered at the mesoregion level. Regressions include mesoregion fixed effects and municipality controls. The predicted migration variable is divided by population in each municipality. Data sources: IBGE data for control variables, COMEX for trade data and SINCRE for migration data. Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1.

Table II.76: **OECD:** Based on σ Elasticity of Substitution at the Extensive Margin

	(1) log(Total Exports)	(2) log(Differentiated Exports)	(3) log(Homogeneous Exports)	(4) log(Total Imports)	(5) log(Differentiated Imports)	(6) log(Homogeneous Imports)
Mercosur Migrants	0.139	0.628	0.229	-0.166	-0.828**	0.193
	(0.295)	(0.388)	(0.322)	(0.351)	(0.395)	(0.250)
Observations	44543	44543	44543	44543	44543	44543
Estimation	IV	IV	IV	IV	IV	IV
Doubling of Mean	0.00334	0.0150	0.00549	-0.00398	-0.0198	0.00463
Municipality Controls	Yes	Yes	Yes	Yes	Yes	Yes
Year F.E	Yes	Yes	Yes	Yes	Yes	Yes
Municipality F.E	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The table shows 2SLS results for subcategories of exports and imports, at the extensive margin, based on the elasticity of substitution σ using the sample of OECD countries. Higher elasticity of substitution means that the product is less differentiated and hence more homogeneous. The dependent variable is of the form $log(x+1)$ and the explanatory variable of immigration is the ratio of immigrants to the population per municipality (in %). Robust standard errors in parentheses clustered at the mesoregion level. Regressions include mesoregion fixed effects and municipality controls. The predicted migration variable is divided by population in each municipality. Data sources: IBGE data for control variables, COMEX for trade data and SINCRE for migration data. Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1.

CHAPTER III

Haitian Immigration and Citizens' Voting Behaviour

Abstract This paper analyses the impacts of immigration on the voting outcomes and electoral participation of natives in Brazil's 2018 elections. This election was marked by the unprecedented rise of a far-right Presidential candidate, Jair Bolsonaro. My analysis focuses on the impacts of one of the largest immigration waves into Brazil, the Haitian immigration, which was triggered by a devastating Earthquake in 2010. I exploit this quasi-natural experiment and I use a variation on the shift-share instrument to predict the immigrants' location decisions. I find that the presence of Haitian immigrants led to an overall average increase in vote shares towards right-wing parties, but had no impact on vote shares for Jair Bolsonaro or Fernando Haddad, the two leading contenders. Haitian immigration also triggered less political participation across different education groups. In particular, a 1 percentage point change in Haitian immigrants relative to the population caused a 0.97 percentage points reduction in turnout ratio, and an increase of 0.89 percentage points in vote shares towards right wing parties, driven by an increase in vote shares towards Presidential candidate Geraldo Alckmin, who was involved in an anti-Haitian immigration controversy. I find that the impact on political participation is present within smaller and poorer municipalities. In contrast, the increase in vote shares to Alckmin seems to be present within small municipalities and those with lower levels of inequality. I analyse potential channels through which immigrants influence voting behaviour and while I rule out mechanisms of natives' fear over public resources, natives' flight (voting with their feet), and changes in candidate

composition, I present suggestive evidence that the change in voter behaviour arises through cultural and economic mechanisms. This paper shows that increases in the ratio of immigrants to the population, even when the migrant community is small relative to the native population, can have important implications on vote shares and can alter the voter composition, affecting the equal representation of the electorate.

1 Introduction

I can't think of a more extremist leader in the history of democratic elections in Latin America who has been elected. - The New York Times, 2018; quoting Scott Mainwaring, Professor at Harvard Kennedy School.

President Jair Bolsonaro won Brazil's 2018 Presidential elections, marking a radical right-wing shift in the country's politics (Londoño and Darlington, 2018). The President engaged in numerous verbal attacks against immigrants, the LGBTQ+ community, women and people of colour, all the while aligning himself to the institutions of religion, family and armed security. Meanwhile, Brazil became a refuge for an unprecedented number of immigrants mainly from Haiti, due to a devastating earthquake, and Venezuela, due to a crippling financial crisis. During the campaign, Bolsonaro was vocal in his anti-immigration attitudes, describing refugees, in a 2016 interview with the New York Times, as "the scum of the world" with directed anger towards Haitian immigrants, accusing them of bringing diseases to the country (Romero, 2016).

This paper focuses on the impact on voter behaviour of the main inflow of immigrants from Haiti in the years leading up to the 2018 election. Political stances on international immigration can be very polarised and politicians can gain more votes by appealing to voters' fears and concerns over this topic. Alesina and Tabellini (2021) have documented a puzzling empirical regularity: anti-immigration sentiment has frequently led to an increase in vote shares towards right-wing parties. This is despite the fact that right-wing policies of less redistribution and tighter fiscal support will typically also impose harm on those voters who are adversely affected by the presence of immigrants. On the other hand, these losses may be balanced by tighter restrictions on immigration, that can help the voter regain a sense of an in-group identity, or face less future taxation which would be needed to fund the increased immediate fiscal support directed towards refugees. Although the literature has focused on such native reactions in response to immigration in richer economies (Dustmann et al 2019; Mayda et al 2019; Tabellini 2020; Steinmayr, 2020), immigrants can also alter voting behaviour for natives in low to middle income countries such as Colombia and the Dominican Republic, but little remains known on the impacts of immigration and the channels in such contexts (Jaupart, 2018; Rozo and Vargas, 2021).

This paper aims to contribute to this literature by asking how, in the context of a middle income country, an increase in immigration can impact political participation

and voting choices of natives.[1] I chose to focus on the 2018 elections due to its polarising nature on immigration issues and to the rather unprecedented rise of the far-right candidate, Jair Bolsonaro, who has been vocal in his anti-immigration stances. This election also occurred after a major, unexpected and novel immigration shock, that led to a wave of Haitian immigration, and after Brazil's government removed the quotas on the humanitarian visas offered to Haitians willing to immigrate into Brazil. Haitian immigration into Brazil was almost negligible before 2010. According to the data, only a hundred or so Haitians resided in Brazil pre-2000.[2] The initial waves of Haitian immigration are considered to be triggered by an exogenous push factor due to the devastating earthquake in 2010 that affected 3 million people, with 250,000 lives lost and 300,000 more injured. About 1.5 million people were internally displaced and made to live in make-shift camps. This earthquake was the most devastating natural disaster to befall Haiti, leaving the country in dire humanitarian need (World Vision, 2021).[3] By the end of 2014, and to respond to the continued and large influx of Haitian migrants, Brazil permitted the unlimited issuance of humanitarian visas for Haitians. This resulted in their arrival in relatively large numbers, and allowed them to relocate and work within the country.

Studying this setting is interesting for four main reasons. First, the study of Haitian immigration provides insights on the impact of new waves of immigrants that had a very meagre presence historically rather than the more established ones that Brazilians are accustomed to having in their country. Second, there existed spatial variation in the locations of the Haitian immigrants that allow me to exploit cross-municipality variation. Third, I have access to electoral outcomes and voter participation at the level of the smallest administrative unit, the municipality. Not only can I study the vote shares allocated to different parties, but I can also study political participation disaggregated by gender, age and education levels. And fourth, I am able to understand and shed light on how immigration interacts with populism in the Latin American context and how this differs from what has been observed elsewhere.

[1] I use the term natives to refer to any citizen of Brazil who is permitted by law to vote. Hence, throughout the paper, I use the terms citizens and natives interchangeably.

[2] This is based on my analysis of the SISMIGRA dataset.

[3] The Haitian immigration started in 2010. However, it wasn't until the late 2013 and early 2014 that Brazil allowed the unlimited increase in the number of humanitarian visas accessible to Haitians wishing to migrate into Brazil (Immigration and Refugee Board of Canada, 2018). An article by IOM UN Migration indicates that it wasn't until September 2015 that the Brazil Visa Application Centre (BVAC) opened its doors to Haitians in Port au Prince. This led to another massive increase in the number of Haitians arriving into the country between 2015 to 2018.

Although the Haitian immigration could be considered a quasi-natural experiment, the locations of migrants within Brazil has been far from random.[4] Therefore, in an attempt to arrive to a causal estimate of the impact of the Haitian immigration, I employ a variation on the traditional shift-share instrument to predict the exogenous element in the location decisions of immigrants, based on the language proximity of Haitian Creole to other languages spoken by earlier immigrant groups. The main assumption behind the validity of the exclusion restriction of this instrument is that historical immigrants located in Brazil due to factors that are independent from current political outcomes. I use this shift share instrumental variable strategy, and discuss different robustness checks on its validity, to predict the number of Haitian immigrants received by each municipality.

I conduct my analysis at the smallest administrative unit in Brazil, the municipality. First, I estimate the impact of a one percentage point increase in Haitian immigration on political participation and outcomes. I study whether the voter turnout ratio differs depending across gender, age and education categories of the electorate. I then provide an analysis of the impact on political ideology groupings of parties and on the individual presidential candidates. Second, I study how the impacts observed vary within different sub-samples of municipalities based on baseline characteristics that are theoretically relevant to political behaviour. Third, I attempt to disentangle the channels of impact that would help explain the results observed. In order to do this, I employ a variety of data sources from *Instituto Brazileiro de Geografia e Estatística*(IBGE, hereafter), the Brazilian 2000 and 2010 Census, and the Latin American Public Opinion Project (LAPOP, 2019) which asks direct questions on voters opinions towards migrants and about their voting behaviour in the 2018 elections. I divide the channels into five categories: economic, welfare-state, natives' flight, candidate composition, and cultural attitudes. I present evidence on the economic channel because natives can react differently to the presence of immigrants based on whether the latter affected their labour market outcomes (Mayda, 2006, Tabellini, 2020), the welfare-state channel because immigration can have impacts on redistribution and fiscal capacity (Facchini and Mayda, 2009), the natives' flight channel because natives can relocate

[4]Initially, the immigration was by land and through a few border cities, that led to overcrowding of some shelters and to Acre government moving some migrants to other cities without the authorities approval leading to nationwide controversies. After the Brazilian government removed the quotas on humanitarian visas in late 2014, the Haitian immigration rose tremendously in number and started to involve air travel to other cities. Haitian immigrants were left to fend for themselves in order to find work, which was legally allowed based on their visa status, and secure their livelihoods.

away from migrant-rich municipalities, the candidate composition channel, reflecting the possibility that immigrants can alter the profile of candidates running for elections, and finally, cultural channel pertaining to natives' perception of the "other" and protectionism over their own identity.

My results show that there were changes in the allocation of votes and in voter turnout in response to Haitian immigration. I find that Haitian immigration led to a change in voter composition due to the differential reduction in the turnout ratio across age and education groups. Further, the results hold true for males, females and across different education groups, but is more pronounced among less educated males. This lower political participation, demonstrated across different disaggregated sections of the electorate, is important and not explored very much in the literature of immigration and political outcomes. Having an unequal representation of the electorate can be problematic for the democratic process, whereby the elected officials no longer reflect an equal representation of the electorate interests (Verba et al. 1995).

The reduction in political participation, as measured by voter turnout, in response to Haitian immigration can be potentially explained through the findings pertaining to cultural attitudes towards immigrants and through the lens of social theories on the lack of political participation. I present individual level explanations to the lack of political participation based on my analysis of the descriptive LAPOP dataset (2019). I hypothesise that the lack of political participation in areas with higher proportion of Haitian migrants occurs because the presence of Haitian immigrants changes the electorate priorities over their concerns. If the electorate doesn't find that the candidates will be able to address these changing concerns over what is considered to be key issues facing their livelihoods, then they will abstain from voting due to uncertainty and apathy. To test this hypothesis, I find that areas that had Haitian immigrants were more likely to think that the biggest problem facing Brazil is discrimination, and those that believed that and were present in Haitian rich municipalities, were less likely to vote. I also find that voters that required a mode of transport to reach the polling station and were also in municipalities with Haitian immigrants, were more likely to abstain from voting than others at the same distance from the polling station. The presence of Haitian immigrants could be increasing the costs associated with voting. Further, based on certain social theories of political participation, the exposure to an *out-group*, the Haitians, could be contributing to a heightened sense of the *in-group* identity, reducing competition within this group and thus creating a demobilising effect (Smith and Mckie, 2000). The presence of Haitians within a municipality could also

change a voter's perceived relative socio-economic ranking within their community. I find that these theories could provide a potential explanation for why the reduction in turnout ratio, in response to an increase in Haitian immigration, is present within smaller and poorer municipalities.

I also find that Haitian immigration led to an increase in vote shares to the right-wing candidate, Geraldo Alckmin, and to a decrease in vote shares to left-leaning parties in the Presidential elections. I find, however, that immigration had no impact on vote shares to Bolsonaro or Haddad, the leading contenders. Alckmin's party, *Partido da Social Democracia Brasileira, PSDB*, represented a prominent opposition party that competed against the long standing left wing *Partido dos Trabalhadores, PT*, before the rise of Bolsonaro, through the *Partido Social Liberal, PSL* party. I observe how this impact varies based on municipality baseline characteristics, where the latter are motivated by political and economic theories of voting behaviour. I find evidence of this impact within municipalities that have lower levels of inequality (defined as having a Gini index less than the median of the sample of municipalities) and within smaller sized municipalities (defined as those with an area at most equal to the median size). Voters in those municipalities with a high concentration of Haitian immigrants voted less for the far-right Jair Bolsonaro. An interesting observation from this analysis is that voters in a sub-sample of municipalities that had a left party win (*Partido dos Trabalhadores, PT*) in 2010 voted significantly less to the PT left wing candidate Haddad and more towards far-right Bolsonaro compared to the same municipalities, in response to an increase in Haitian immigration. Those impacts point to a *natives' backlash* (Tabellini, 2020; Jaupart, 2018; Rozo and Vargas, 2021), in response to the presence of Haitian immigrants, towards right-wing candidates.

I present suggestive evidence that the significant increase in vote shares to Alckmin is driven mainly by voters switching their preferences away from other candidates, and not solely driven by a reduction in political participation. I then proceed to study possible channels to explain the impact. I rule out the channels of natives' flight and changes in the candidate composition as explanations for the reported results. I also find evidence suggesting that the welfare-state channel (i.e. the possibility that natives might be fearful over the sharing of public resources), was not a main channel for the results. My evidence indicates that an increase in Haitian immigration relative to municipality population had no impact on taxation nor municipal social transfers and expenditures, but had a rather negative impact on agriculture expenditure. My analysis hence suggests that cultural and economic channels alongside the candidates'

ideological positions provide more plausible mechanisms for the observed results. My findings are consistent with theories on political participation and voting behaviour in response to immigration.

First, cultural related attitudes towards Haitian immigrants could be a potential explanation for the observed impacts. For instance, Alckmin was painted with an anti-immigration brush at the onset of the Haitian crisis. In his earlier capacity as governor of São Paulo, Alckmin got directly involved in a nation-wide political controversy speaking out against the arrival of approximately 800 Haitian immigrants at the start of the immigration (Cogo and Silva, 2016).Through my analysis on cultural attitudes, I show that it is more likely that voters choose Alckmin or Bolsonaro over abstaining from the elections as the ratio of Haitian immigrants to the population increases. However, I find evidence that supports the *Contact Hypothesis* (Alesina and Tabellini, 2021) when studying the impact on vote shares within municipalities where Haitian immigrants make up a larger share of the population. Within this sub-sample, I find that there is a reduction in vote shares towards Bolsonaro. These results lend support to the contact hypothesis, which states that exposure to a different cultural group can begin to create cohesion among the different groups that make up the society, and hence one would expect a reduction in vote shares to anti-migration candidates. This finding is consistent with other results in the literature (Steinmeyer, 2020). And second, my evidence is indicative of a very small but positive economic impact on municipal outcomes due to the presence of immigrants. An increase in the ratio of Haitian immigrants to the population led to positive impacts on the value added in services and industry as well as the number of companies. Voters could be choosing the candidate that has promise in delivering economic prosperity that would allow the assimilation of the new immigrant labour force. Based on the analysis of ideological positions across a multitude of topics (dataset based on Wiesehomeier, at al., 2021), Alckmin is a business-friendly figure who opposes all forms of state ownership of businesses and industry and promotes the deregulation of markets, more than Bolsonaro does (Wiesehomeier, at al., 2021).

I contribute to the literature in the following ways. First of all, I complement the literature in understanding how immigration impacts political behaviour in a middle income country and I shed light on the channels at play (Jaupart, 2018; Rozo and Vargas, 2021; Dustmann et al. 2019; Mayda et al 2019; Tabellini 2020; Steinmayr, 2020). Further, I study the context of Brazil, the fourth largest democracy in the world, at a unique moment in its history that marks the rise of a far-right presidential

candidate after years of dominance by the left-wing "Workers Party". Second, I present evidence on the impact on electoral participation, in terms of turnout ratio, that is disaggregated by education, gender and age. This level of disaggregate analysis has not been given much attention in the economic literature tackling this question. Through this analysis, I study how immigration can impact the voter composition and rationalise the findings within social and political theories of voter turnout (Campbell, 2013; Mutz, 2006; Solt, 2010; Kernet al., 2015; Dalton and Klingemann, 2011; Jost, 2021; Soss and Jacobs, 2009; Smith and Mickie, 2000). Understanding the impacts on voter turnout and vote shares is important because the former affects equal representation in a democratic system and the latter affects political outcomes and public policy reforms that have long-term welfare consequences. Although this goes beyond the scope of this version of this paper, my future work will look into the repercussions from the changes in the electoral behaviour of natives on local social and economic outcomes. Third, my research also focuses on an episode of south-south immigration. In particular, I study the large and sudden immigration, from Haiti, which was triggered by an environmental disaster, as opposed to the more organised and gradual economic migrant inflows (Ruiz and Vargas-Silva, 2018). These migration episodes are likely to continue to happen due to climate change, political upheavals and economic crises, and therefore quantifying and rationalising their impacts on the host community is important to guide the policy space on the absorption of the migrants. Those immigrants are different from other voluntary immigrants in that they often arrive without sufficient economic means, are traumatised by conflict, and face a high level of uncertainty regarding their mobility and length of stay in the host community. Hence, the impact that can arise due to the sudden arrival of a wave of immigrants may be very different from that due to gradual, and voluntary immigration. Fourth, Brazil is a country that had to innovate on its immigration laws to allow for the legalisation of humanitarian immigrants (Pacifico et al., 2015). The immigration structure and laws, alongside constrained fiscal resources, mean that the assimilation and effects of immigration waves that are triggered by environmental or humanitarian causes will be different in this context than that of a developed country. And finally, I expand the evidence on the channels of impact by providing descriptive analysis on the Latin American Public Opinion Survey (2019) to assess cultural attitudes towards immigration as it pertains to political outcomes and participation (Alesina and Tabellini, 2021; Tabellini, 2020).

This paper proceeds as follows. Section 2 presents a literature review on the recent empirical and theoretical evidence pertaining to the links of immigration to electoral

outcomes. Section 3 discusses the political landscape in 2018 and the historical background on Haitian immigration into Brazil. Section 4 discusses the data used in this analysis and their respective sources. Section 5 discusses the estimation methodology, and elaborates on the instrumental variable approach, and the different robustness checks and assumptions. Sections 6 presents the results from the analysis, and it is divided into five subsections: section 6.1 presents results on the voter turnout ratio, section 6.2 presents the results on the electoral results in terms of vote shares, 6.3 presents analysis on how the results vary depending on municipality characteristics, section 6.4 discusses potential channels of impact, and section 6.5 presents an integrated interpretation of the results. Section 7 concludes.

2 Literature Review and Channels of Impact

My work speaks to and contributes to two strands of literature. A first strand of the literature relates to the impact of immigration on political outcomes of the host economy and the underlying mechanisms, and the second strand of the literature relates to the determinants of political participation. In what follows, I summarise the recent literature that my paper speaks to, organised within the context of the various channels of impacts.

2.1 Political Participation

The act of voting is an expression of preference (Downs, 1957 as cited in Sanders 2001). Abstention in of itself can be considered a political choice (Alvarez and Nagler 1995, 1998). Studying how immigrants affect political participation is important because such an impact can alter the voter composition and representation of the electorate. The determinants of political participation have been explored extensively in the political science and sociology literatures. For an in depth review of the literature on the determinants of voter turnout, see Stochemer (2017). In what follows I discuss some of the factors that would impact voter turnout and the role of immigration through which it can heighten or lessen the influence of those factors. I abstract from a discussion on the importance of institutional and national level factors as determinants of voter turnout as those would be less relevant for my context. The reason is that such factors are common across all municipalities, whereas my analysis exploits the spatial

variation of Haitian immigrants within the country.[5]

2.1.1 Individual Channels With a Social Dimension

Multiple individual-level characteristics help explain voter turnout. Those characteristics operate through a social context and are mediated within the social network of the voter (Campbell, 2013). Sanders (2001) proposes that individuals turn out to vote for their preferences only if the utility difference between a pair of candidates exceeds a certain threshold. Further, voters make their choices with a certain degree of uncertainty. This uncertainty can arise because information on candidates' positions is costly, there are cognitive limitations on the part of the voter, and ambiguity on the part of the candidate (Sanders, 2001). Predicting the probability that a voter will turn out to vote based on the difference between expected benefits and costs of voting has its links in the rational voter hypothesis (Downs, 1957). However, from this assumption rises the "the paradox of voting": the expected benefits for a voter depend on the probability that their vote is decisive, which can be approximately zero for most national elections across the world (Kaniovski and Mueller, 2006). To resolve this paradox, other explanations on the determinants of political participation have been suggested emphasising the role of the voter's social context. For instance, people may vote out of a strongly held sense of civic duty.

Campbell (2013) highlights the importance of individual characteristics in terms of the voter's level of education, religious attendance, mobilisation, political knowledge, strength of preferences, as additional predictors of voter turnout and ones that operate within the specific social context of the voter. Individual education relative to others within the same social sphere is a powerful predictor of political participation (Campbell, 2013). In a similar vein, individuals who frequently attend religious services tend to have higher levels of political participation. Putnam and Campbell (2011) (as cited in Campbell, 2013) investigate why this is the case within the American context. The authors find that the key aspect driving these results is the social networks formed within the church. Interestingly they find that many non-believers with church friends tend to have higher levels of political participation than socially isolated believers. Mobilisation is another determinant of political participation, and it refers to being

[5]Among those factors are increases in party competition through variation in the number of parties running for election (Capron and Kruseman, 1988); the electoral system type, importance of the election, voting age, registration laws

invited in a way or another to participate in a political activity. However, the relationship between who is inviting and who is receiving the invitation matters in the effectiveness of the message. Moreover, mobilisation efforts can be more concentrated in areas that are more homogeneous in a certain dimension and where they are likely to lead to higher effectiveness. Political knowledge is yet another important determinant of turnout, and although it can be framed as an individual characteristic, it is often influenced by the environment that the voter is in. Besides formal education, the voter can self select into groups where they can learn more about politics. To overcome the self selection issue, field experiments shed light on a causal impact. For instance, Klofstad (2011) (as cited in Campbell, 2013) finds through a field experiment that uses lottery to assign students to dorm rooms, that those who had more political discussions were more likely to turn out for elections.

In a similar vein, the strength of preferences that voters hold can determine their willingness to show up to the voting booths. What's interesting is that voters' preferences can also be mediated through the level of diversity within a certain social context. Conviction can be reinforced within a community of like minded people, but can also turn into ambivalence in settings where there is opposing political opinions. This ambivalence dampens political participation (Campbell, 2013; Mutz, 2006). However, establishing causality in this context is difficult because of selection bias and reverse causality.

2.1.2 Socio-economic Channel

Immigrants also alter the ethnic and racial composition of the electorate within a certain municipality. Therefore it is important to consider social theories in the explanation of voter turnout that go beyond individual choices. Jost (2021) argues that it is important to make a differentiation between variation in affluence *within* a certain area as opposed to those *across* spatial areas. Soss and Jacobs (2009) argue that large differences between spatial units can produce "distinct and homogeneous clusters of unequal social privilege and political influence, intensifying the social and political isolation and polarisation of neighbourhoods". So if immigration alters affluence both within and across municipalities, it can increase political isolation, leading to lower electoral participation. Smith and Mckie (2000) argue that being exposed to an "out" group solidifies the identity of the "in" group. In areas that became heterogeneous because of the presence of immigrants, a sense of an in group identity can become

heightened across all factions of the society. This might create a sense of homogeneity among the natives, reducing the competition within this group, resulting in a demobilising effect.

There have been other papers that study the change in turnout ratio as a consequence of ethnic heterogeneity of the community. Belletini et al.(2016) explore, in an Italian context, the relationship between political participation and ethnic heterogeneity at the neighbourhood level. They find that political participation is less in ethnically heterogeneous neighbourhoods. This impact is an indirect one as the authors examine the effect of immigrants, who are not entitled to vote, on the turnout ratio of citizens, who are. Other studies have confirmed this result in different contexts. Barber and Imai (2013) show through a rich individual level dataset of voter records in three US cities, that an increase in the out-group proportion, defined in terms of political affiliation or race, depresses the probability of political participation. Fieldhouse and Cutts (2008) take the British General Election of 2001 as a case study to investigate this question and find that turnout increases among members of the minority group, Asian in this case, as the size of the latter increases in a certain municipality. However, the turnout of their non-Asian neighbours from other (majority) groups decreases.

These results can be rationalised through the psychological theory of voter empowerment, which states that voters who reside in heterogeneous localities turn out less frequently because of their natural tendency to avoid conflict or due to the moderating effect of being exposed to opposing views. But this observation is also consistent with other channels; for example, the mobilisation theories might apply if mobilisation campaigns target more homogeneous communities to maximise the efficacy of their efforts (Barber and Imai, 2013).

2.2 Voting Outcomes

Immigration can impact voting outcomes and behaviour in two main ways. First, immigrants who become naturalised can directly affect the voting outcomes. Second, immigrants can indirectly affect voting outcomes through their impact on the utility of native voters or through altering the ideological agenda and policy. In this paper, I focus on indirect impacts of immigrants on electoral outcomes. In Brazil, only native and naturalised citizens that are permitted legally to vote in the elections. The majority of Haitian immigrants were given temporary or permanent visas or sought asylum, and

so would have not been allowed to vote in the 2018 Presidential elections. Immigrants can impact the voting behaviour of natives indirectly through many different channels. These channels can be broadly characterised into three categories: economic, welfare-state, and cultural.

2.2.1 Economic Channel

The economic channel through which immigration affects electoral outcomes was characterised by Benhabib (1996), who studies how immigration policy that determines the level of skill required by immigrants could be determined under majority voting as natives choose policies to maximise their well-being. Key to the analysis of the labour market channel is that immigrants are heterogeneous in terms of their skill level. Therefore their influx into a country, alters the skill composition of the electorate (in the long term when immigrants gain the right to vote) and changes the relative proportions of low to high skilled labour. The theory suggests voters are likely to favour pro-immigration policy when the immigrants possess a skill level that is complementary to these voters, and tend to oppose immigration when the immigrants and the natives are substitutes in the labour market (Ortega, 2005). Labour market competition between immigrants and natives can translate into reductions in wages and higher unemployment for native-born populations. In these cases, immigration is likely to lead to a shift towards political candidates who support immigration only of people whose skills are complementary to those of the median voter, and who will correspondingly oppose all immigration of those with other skill types, (Mayda, 2006; Ortega and Polavieja, 2012).

However, the direction of the economic impact and who wins and loses is far from straightforward. The empirical literature has found that, depending on the context and methodology used, immigrants can either have a negative short lived impact, no impact, or a positive impact (Alesina and Tabellini, 2021). Some papers have found a clear negative impact of immigration on wages. For example, Dustmann et al. (2017) find that the immigration shock of Czech workers in Eastern Germany had significant negative effects on the employment of low-skilled natives. Monras (2020) studied the impact of Mexican migrants across US cities and found that immigration impacted negatively wages of natives in low-skilled occupations. However, this impact was concentrated in the short- term and dissipated over the subsequent five years as the market adjusted. Others have found that there is no significant labour market

impact on natives due to immigration. Clemens et al. (2018) studied the context of the *Bracero* program in the United States; the authors found that the exclusion of 500,000 Mexican workers had no labour market impacts on US employees because technological change endogenously responded. And finally, studies have also presented evidence lending support to the idea that immigrants can have a positive impact on the labour market even among the lower skilled (Peri and Sparber, 2009; Battisti et al. 2018).

2.2.2 Welfare-state Channel

The second channel of impact through which immigrants can impact natives' political behaviour is due to welfare concerns related to redistribution and fiscal considerations. For example, natives might fear that the presence of low-skill immigrants, especially those that require humanitarian action, will put additional stress on local fiscal resources, potentially leading to an increase in taxation. However, when immigrants are skilled, they might result in a benefit for the welfare-state. In either case, the presence of immigrants will likely result in redistribution which in turn could impact voters views towards immigrants. Facchini and Mayda (2009) develop a theoretical framework that links welfare-state actions to the skill levels of a group of immigrants and thus leads to changes in the political dynamics and political discourse on immigration policy.

Burgoon and Rooduijn (2020) show that anti-immigration attitudes can generate competing impacts in the politics of welfare-state generosity. On one side, natives who have such attitudes may prefer political candidates that who promise to limit welfare redistribution to shield immigrants from receiving transfers. On the other hand, however, natives may be supportive of greater welfare redistribution if they perceive the presence of immigrants as a threat to their economic wellbeing. Burgoon and Rooduijn (2020) argue that whether or not natives support welfare redistribution is context specific and depends on the size of the foreign born share, the size of the welfare state and the eligibility of immigrants to social services. Where the latter two factors are high, natives tend to support lower welfare redistribution. Immigration can also be impacted by the type of the welfare state whereby the latter can affect the incentives to immigrate in the first place or the longevity of stay in the destination country (Borjas, 1999).

Alesina et al. (2022) show, through large-scale surveys in six countries, that not only do natives over-estimate the size of the immigrant community and under-estimate their education and employment levels, relative to theirs, but also are less likely to support redistribution when thinking about immigration beforehand. Their views on redistribution are less malleable to the true facts of the migrant community and are rather shaped by narratives.

2.2.3 Cultural Channel

The third channel through which immigrants can impact voting behaviour is related to cultural and identity explanations. Card et al. (2012) points to a channel where natives can develop anti-immigration sentiment if they feel that immigrants are threatening the "compositional amenities" from common resources, such as neighbourhoods, schools or workplaces. Interestingly, the authors find that compositional concerns are two to five times more important in explaining the variation in anti-immigration attitudes than labour market concerns (Card et al., 2012). Further, the group-conflict theory suggests that natives might foster hostility towards immigrants if they perceive the latter group to threaten their identity and culture. Tabellini (2020) studies the impact of immigration across US cities from 1910 to 1930. The author finds that the increased support for anti-immigration parties in the US congress, had roots in the cultural differences between immigrants and natives, as opposed to economic roots. However, while it is conventional wisdom that immigrants will trigger a negative reaction among the electorate, the Contact Hypothesis states otherwise. In fact, the hypothesis states that inter-group contact can have positive impacts on the cohesion across groups, especially if the minority group is not seen as a threat to the identity of the larger group. There is also support for this hypothesis, empirically. Steinmeyer (2020) shows that voters in Austrian municipalities that had a prolonged exposure to refugees voted less for anti-immigration candidates as opposed to voters in areas that had only a fleeting response to the refugees.

Given these multiple and ambiguous effects, the political impact of immigration cannot be discerned a priori. The conflicting evidence in the literature also suggests that this impact is very much context specific: not only does it depend on the country in question, but also the political leadership and the nature of the right-wing or left-wing populism. It also depends on the type of immigrant, whether they are complements or

substitutes in the labour market and how culturally distant they are from the native population.

Although the impact of immigration on natives' electoral outcomes has received substantial of attention in a developed country context, the literature studying this impact in low and middle income countries is still growing. Hence, I find my paper to be closest to work done by Jaupart (2018) who studies the impact of the Haitian immigration on electoral outcomes in the Dominican Republic and to Rozo and Vargas (2021) who study the impact of the Venezuelan refugee crisis on electoral outcomes in Colombia.

Jaupart (2018) finds a positive and robust relationship between the impact of Haitian immigrants on the vote shares of the right wing party that has traditionally been opposed to immigration. He also finds a negative impact of immigration on the vote shares of parties that are more immigration friendly. Jaupart (2018) complements this analysis by shedding light on cultural channels such as individual concerns over citizenship, political competition and cultural identity. Although my article also studies ta migration influx from Haiti, , the context of Brazil remains very different to that of the Dominican republic. For one, the historical relationship, between the Dominican Republic and Haiti is rich. Haiti occupied the Dominican Republic for over 20 years after 1844 and until this day, the Dominican Independence Day represents freedom from Haiti, as opposed to Spain or UK. This fostered an attitude of resentment towards Haitians, and the Dominican identity in recent years under the rule of Joaquin Balaguer was built on its differentiation from the Haitian one (Jaupart, 2018). Further, the Haitian immigration in terms of its volume, timing and distribution differs widely in the Dominican Republic from that in Brazil. While it is a rather recent phenomena in Brazil (gained momentum after 2010), the immigration has deeper roots in the Dominican Republic where Haitians represented 3.3% of the population and 80.7% of the foreign born population (Jaupart, 20018).

Rozo and Vargas (2021) study the influx of 1.3 million Venezuelan immigrants into Colombia as a consequence of the Venezuelan humanitarian crisis. The authors find that larger immigration concentration leads to an increase in voter turnout ratio and an increase in votes towards right-wing parties. The authors point to two main channels of impact that could be at play. The first is a novel channel that the authors term as *electoral misinformation*. This channel suggests that political parties in Colombia campaign against the opposing party if the latter's ideological position is similar to that of the regime that triggered the forced immigration in the first place. The authors find

support for this channel in their context where municipalities that historically faced left-wing illegal insurgencies, had a much more pronounced increase on right-wing vote shares. The other channels relate to the economic impacts of immigration, where they find that Venezuelan immigration increased unemployment.

3 Background

The context I study is rich and evolving, as the Haitian immigration continues into Brazil and as the political landscape develops. I present in this section a brief description of the Haitian immigration to Brazil and of the presidential political landscape in Brazil.

3.1 Haitian Immigration

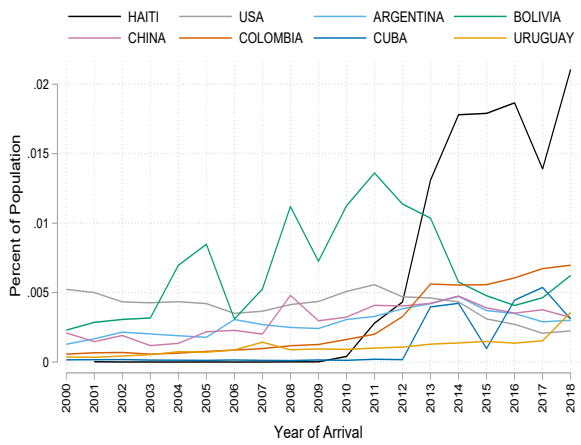
There was little connection between Brazil and Haiti until 2004. In that year, Haiti experienced a major internal conflict that resulted in the removal of President Jean-Bertrand Aristide from power. Following these events, the United Nations sent a peacekeeping mission to maintain peace; this mission was known in French as *Mission des Nations Unies pour la stabilisation en Haiti (MINUSTAH)*. Brazil aspired to become more powerful in the international arena during that period and agreed to lead the military component of MINUSTAH (i.e. led by the Brazilian Army with a Brazilian force commander) in order to support those ambitions (Müller and Steinke, 2018). MINUSTAH set a precedent as the first mission in the region to be led by another Latin American country. Haiti has been traditionally a country of emigration. According to some analysis, Haiti has the world's highest high-skilled emigration rate (Jadotte, 2012).

However, there were few Haitians living in Brazil until 2010. In that year, after a long period of the Brazilian Army leading MINUSTAH, Haiti experienced a devastating earthquake. Death toll estimates are in the hundreds of thousands (Braga, 2010). The earthquake also shrunk Haiti's economy , with estimates of GDP losses greater than 10% (Best and Burke, 2019; Cavallo et al., 2010). The existing link between Brazil and Haiti because of MINUSTAH, together with the devastating economic and social consequences of the 2010 Earthquake, and the promise of a more economically prosperous future in Brazil, led to a large inflow of Haitians to Brazil. I report the yearly inflow of Haitians as a proportion of the yearly total population in Brazil in figure

III.1(a). To reach Brazil, Haitians typically had to travel to the Dominican Republic (land or air), fly to Ecuador and then travel by land across Peru to enter Brazil in one of the land borders with this country. Haitian immigrants then normally follow the path towards the cities of Assis Brazil, Epitaciolaândia and Basilia or Tabatinga. There are cases where the immigrants also arrived through Corumb'a, Mato Grosso, Grosso do Sul or by airports to major cities in Brazil like São Paulo (Cogo, 2014). While it was common for Haitians to enter Brazil via the Peruvian border, most eventually settled in other regions of the country (Cavalcanti et al., 2017). Importantly, as shown in Figure III.1(b), these crossing points into Brazil were mainly: Tabatinga (Amazonas state) and then on to Manaus or through Brasiléia (Acre state) or Rio Branco, before dispersing into different municipalities throughout the country later on.

Figure III.1: The Migratory Flow and Entry Points of Haitians

(a) The Flow of Immigrants into Brazil by Nationality



(b) Route and Main Entry Points of Haitian Migrants into Brazil

Note: Figure III.1(a) depicts the flows on immigrants into Brazil since 2000 as a percent of the population. The flows represent the summation, by country of origin, of all immigrants coming into Brazil in that year. To make the figure readable, the figure depicts only countries of origin with at least 4,000 immigrants across all years. The figure also omits the Venezuelan immigration into Brazil because of its large volume. Source: SISMIGRA dataset. Figure III.1(b) shows the entry routes of Haitians into Brazil. This is an adapted figure from the original figure present in Rawlinson et al. (2014).

By 2014, Haiti began the source country with the largest number of migrants in Brazil's formal labour market, due in part to a few firms attempting to hire Haitians from the refugee camp in Acre in the early years of the crisis (Cavalcanti, Oliveira

and Tonhati, 2014; Beekma, 2015).[6] I present in table III.1 the demographic profile of Haitian immigrants who arrived to Brazil from 2010 to 2018. Among the immigrants who arrived into Brazil, 64.9% were male and only 35% were females. In terms of the age groups of the immigrants, the presence of children and elderly people is hardly noticeable.[7]Haitian immigrants were almost entirely individuals in working age, from 25 to 40 years and basically formed by single (unmarried) individuals (73.8%). The migration data also provides information on the profession of the immigrant.[8] For those immigrants reporting an occupation, I classify them into four broad categories: low skill, high skill, student and dependent. A student is any immigrant who reported "student" as their profession, and a dependent is anyone who is reported as a minor or a dependent on a temporary or permanent visa holder. The low skill category encompasses professions such as: "Agriculture worker", "Barber", "Driver", "Dry Cleaner", "Packaging", etc. While the high skill category encompasses professions such as "Cashier", "Engineer", "Nurse, "Scientist", "Technical Profession" etc. I report that 34.3% of the Haitian workers are classified as low skill while only 11.4% are classified as high skilled.

My dataset constitutes of 125,000 Haitian immigrants who entered Brazil between 2010 and 2018.[9] During this period, Haitian immigrants were dispersed across Brazil. Prior to 2000, there were only 118 Haitian immigrants living across the entirety of Brazil, and these were scattered across 48 municipalities. By 2018, there were over one hundred thousand Haitian immigrants, living in 328 different municipalities (see figure III.2 that visually depicts the spatial distribution of Haitian immigrants across Brazil).[10] I provide statistics on the locations of these municipalities, and median size (in km^2) in table III.28 in the appendix. I find that initially, 17.6% of the immigrants were present in the North of Brazil, a proportion that declined in the years after the

[6]The camp was established in 2010 in Brasiléia, in Acre state, but was later forced to shutdown in April 2014 due to severe overcrowding and severe floods from a nearby river, the Madiera, that prevented the availability of food and supplies. Only 2,000 of those refugees were transferred to another nearby camp in the same state, where the rest of the migrants were left to fend for themselves. Haitian migrants waited there to receive their work permits in order to be able to disperse throughout the country finding work through their networks or word of mouth on available jobs in the South (Borgon, 2014; Davies, 2014; Harumi Miura, 2014). As of 2015, Acre stopped being the main route of entry for Haitians as the Brazilian government began issuing visas to Haitians more abundantly (Sousa, 2015)

[7]The following number rely on both the STI-MAR and the SISMIGRA datasets on immigration provided by OBMigra.

[8]But I find this measure to be very noisy with 35.6% of Haitians reporting a missing profession.

[9]This number (125,180) is approximately equal to the number reported by the Migration Policy Institute (MGI, 2020).

[10]Based on the analysis of STI-MAR and SISMIGRA datasets accessed on June 2022.

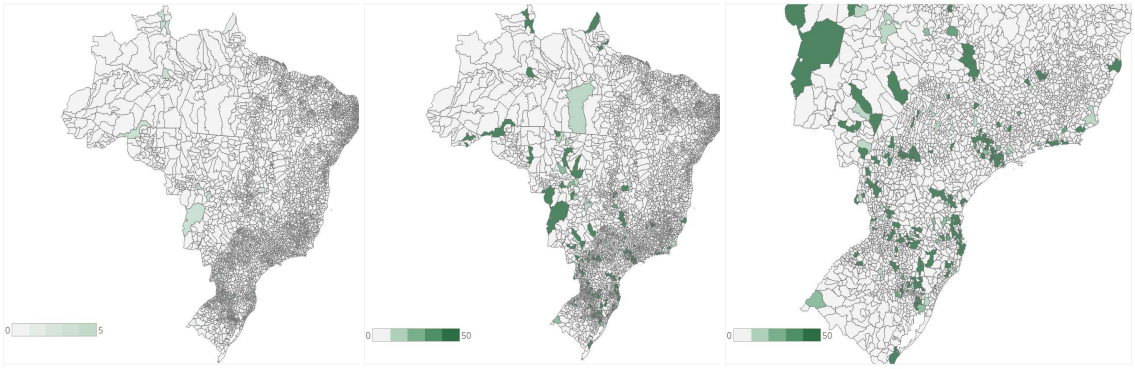
Earthquake to 4.4%. This is understandable considering that the main points of entry into Brazil were in the Northern region and as can be seen from the table, the majority of Haitian immigrants settled ultimately in the wealthier South and South-East regions of Brazil. Although they seem to have initially settled in larger municipalities (median area size of 6,500 km^2 and a population of 1,800,000 and 1,700,000), they later moved to smaller municipalities with approximately median area size of 625 km^2 and a population just north of 300,000 individuals.

The total number of Haitians makes them the second largest migrant nationality in Brazil, after Venezuela. Even so, however, the total size of the Haitian immigrant population is small relative to the overall population of Brazil. Even at the municipality level, the spatial distribution of Haitian immigrants implies that there are few municipalities where Haitians represent even one percent of the population; see the cumulative distribution of immigrants in figure III.17(b) the appendix. In magnitude, Haitian migration into Brazil is dwarfed by other recent episodes, such as the influx of Syrians into Lebanon and Jordan. With such relatively small presence of Haitians in Brazil, it is to some extent surprising to find any impacts at all on electoral outcomes. I hypothesise that the small numbers suggest that electoral effects are unlikely to be mainly driven through large labour market impacts. Instead, political impacts seem more likely related to media representations and other perceptions. Esses and Medianu (2013) discuss how media portrayal of immigrants paints them with a dehumanising brush that suggests that they spread diseases or are terrorists in disguise. In a similar vein, Abid et al. (2017) finds that host countries use the term "burden" in the media to describe Syrian refugees as a wide spread "problem" affecting the economy and environment. Ohme (2020) points to the role of frequent social media consumption which can strengthen existing attitudes making those who become more rigid in their opposition to immigration issues less likely to be politically active. Further, Steinmeyer (2020) shows that municipalities in Austria that were merely crossing points for migrants with a fleeting interaction with the natives had a much stronger anti-migration backlash than those that actually had migrants residing in them.

When the earthquake crisis hit Haiti triggering migration flows into Brazil between January 2010 and February 2011, immigrants applied for Asylum status. The National Committee for Refugees (Comitê Nacional para os Refugiados, CONARE) did not consider that immigrants fleeing their country due to environmental crises should

Figure III.2: The Distribution of Haitian Immigrants

(a) Haitian Migrant Flows till 2000 (b) Haitian Migrant Flows 2010-2018 (c) Haitian Migrant Flows 2010-2018



Note: The figure depicts The distribution of Haitian immigrants across the Brazilian Municipalities before and up until 2009, and between 2010 and 2019. Figure depicts flows of immigrants based on year of entry. The flows are a summation of both the Venezuelan and Haitian immigrants and are depicted in absolute terms, not as a proportion of each municipality's population. Source: STI-MAR and SISMIGRA dataset.

be given refugee status (Jubilut et al., 2016).[11] However, the National Immigration Council of Brazil (Conselho Nacional de Imigracão, CNIg), considered the asylum claims by Haitians and allowed them to stay for humanitarian reasons. This was implemented under the Resolution No. 08/06 (Nieto, 2014; de Souza Rodrigues 2016). The CNIg examined the requests, and it was under its discretion to grant the humanitarian visas. By 2014, 600 asylum applicants were offered the authorisation to remain in the country under a humanitarian visa (IMDH, 2014).

Later, CNIg adopted the Normative Resolution No. 97 which limited the number of humanitarian visas that can be offered to 1,200 per year, which translated to only a 100 per month. Those visas excluded family reunification requests and were issued only in the Consulate of Brazil in Port-au Prince (IPPDH and IOM, 2017; Nieto, 2014). This law failed to curb the Haitian influx into Brazil to the extent that many Haitians got stuck in Inãpari, a city in Peru bordering Brazil, and the state of Acre in Brazil had to establish camps for the Haitians. This led to revoking Resolution No. 97 and implementing instead resolution No. 102 in 2014 that eliminated the limit on the Haitian immigrants coming into the country, and allowed the visas to be issued in Brazilian consulates in Ecuador, Bolivia, and the Dominican Republic (IPPDH and IOM, 2017). Further, the Brazilian government created the Humanitarian Visa Ap-

[11]CONARE stands for the National Committee for Refugees. CONARE is linked to the Ministry of Justice and is made up of representatives of the following Ministries: Justice, Foreign Affairs, Health, Education, as well as the Federal Police and civil society organisations dedicated to assistance, local integration and refugee protection in Brazil. UNHCR and the Office of the Public Defender of the Union have a seat in CONARE with a voice but not a right to vote (UNHCR).

plication Centre in Port-au-Prince in Haiti in a partnership between the Ministry of Foreign Affairs (MRE) and the International Organization for Migration (IOM) to expand the approvals of humanitarian visas in 2015. Finally, in April 2018, Brazil adopted the Inter-ministerial Order No.10, which sets out procedures to evaluate and offer temporary and residence permits based on humanitarian reasons for Haitian nationals and stateless people in Haiti. Those visas were only allowed to be issued in the Brazilian embassy in Port au Prince (Immigration and Refugee Board of Canada, 2018).

It is worth noting that Brazil in 2017 received another major immigration episode from Venezuela. Due to the political turmoil and socio-economic instability, Venezuela faced one of the largest external displacement crisis in the history of Latin America. However, I do not study the impact of the Venezuelan immigration on political outcomes in this paper. I focus on the prolonged wave of Haitian immigrants, who have known residential areas in Brazil based on the data I use in this study As not to bias the results on the Haitian immigration, I exclude the two main municipalities of entry of Venezuelan migrants from the analysis, Boa Vista and Pacaraima.

Table III.1: Migrant Characteristics by Nationality Across All Years

	Migrants					
	All Other		Haitians		Total	
	Total	%	Total	%	Total	%
Sex						
Female	282,853	36.1%	43,870	35.0%	326,723	36.0%
Male	500,579	63.9%	81,303	64.9%	581,882	64.0%
Missing	57	0.0%	7	0.0%	64	0.0%
Total	783,489	100.0%	125,180	100.0%	908,669	100.0%
Skill Level of Profession						
Low Skill	116,455	14.9%	42,972	34.3%	159,427	17.5%
High Skill	262,078	33.5%	14,280	11.4%	276,358	30.4%
Student	102,906	13.1%	7,598	6.1%	110,504	12.2%
Dependent	23,443	3.0%	1,755	1.4%	25,198	2.8%
No Occupation	27,644	3.5%	13,989	11.2%	41,633	4.6%
Total	783,489	100.0%	125,180	100.0%	908,669	100.0%
Age Category						
0 - 15 age	50,745	7.7%	6,725	5.9%	57,470	7.4%
15 - 25 age	180,125	27.3%	23,828	20.9%	203,953	26.4%
25 - 40 age	274,826	41.7%	70,333	61.8%	345,159	44.7%
40 - 65 age	144,785	22.0%	12,821	11.3%	157,606	20.4%
65 - age	8,420	1.3%	127	0.1%	8,547	1.1%
Missing	43	0.0%	6	0.0%	49	0.0%
Total	658,944	100.0%	113,840	100.0%	772,784	100.0%
Civil Status						
Single	533,737	68.1%	92,391	73.8%	626,128	68.9%
Married	219,831	28.1%	27,214	21.7%	247,045	27.2%
Separated/Divorced	3,652	0.5%	112	0.1%	3,764	0.4%
Widower	11,140	1.4%	152	0.1%	11,292	1.2%
Other	15,129	1.9%	5,311	4.2%	20,440	2.2%
Total	783,489	100.0%	125,180	100.0%	908,669	100.0%
Visa Type						
Asylum	124,545	15.9%	11,340	9.1%	135,885	15.0%
Border	517	0.1%	0	0.0%	517	0.1%
Other	3,503	0.4%	10	0.0%	3,513	0.4%
Permanent	147,580	18.8%	75,191	60.1%	222,771	24.5%
Provisional	806	0.1%	27	0.0%	833	0.1%
Resident	33,495	4.3%	18,976	15.2%	52,471	5.8%
Temporary	473,040	60.4%	19,636	15.7%	492,676	54.2%
Total	783,486	100.0%	125,180	100.0%	908,666	100.0%

Note: The table shows the distribution of immigrants across key individual characteristics: sex, skill level based on their indicated profession, age category, civil status and visa type. The table shows the total numbers of flows of immigrants and the percentage of immigrants across the categories accumulated between the years of 2010 to 2018. Source: SISMIGRA dataset.

3.2 The Brazilian Political System and The Rise of the Far-Right

Brazil is politically organised as a Federal Republic, formed by the union of 27 states (one is the Federal District, Brasilia) and 5,570 municipalities (Ramirez, 2021). Approximately, double the area size of the European Union and half its population, Brazil is the biggest democracy in Latin America (Ramirez, 2021). The Brazilian constitution upholds the principle of the separation of powers into legislative, executive and judiciary. The executive branch is led by the president and vice president of the republic. They are both elected every four years by absolute majority through a two-round system. If the president obtains absolute majority in the first round of the elections, then they are considered the winner. But if they don't, then the two leading candidates compete in the second round of the elections.

Brazil's legislative power is held by the National Congress. The National Congress is a bicameral parliament that is composed of two key institutions: the Chamber of Deputies and the Federal Senate. The Senate is elected every four years-alternating $\frac{1}{3}$ or $\frac{2}{3}$ of seats. It represents the 26 states and the Federal District. Each state and the Federal District has a representation of three Senators, who are elected by popular ballot for a term of eight years. Every four years, renewal of either one third or two-thirds of the Senate and of the delegations of the States and the Federal District takes place. Senators are directly elected by the residents of each state. The Chamber of Deputies is elected for a four-year term by a system of proportional representation. Seats are allotted proportionally according to each state's population, with each state eligible for a minimum of 8 seats (least populous) and a maximum of 70 seats (most populous).[12]

[12]This system grants a disproportionate share of political power to the states of the Northeast and North and severely under-represents the heavily populated state of São Paulo. Unlike the Senate, the whole of the Chamber of Deputies is renewed every four years.

Table III.2: The Political Federal System in Brazil

	Federal	State	Local
Territory	All of Brazil	26 States + 1 Federal District	5,570 Municipalities
Executive	1 President, 1 Vice-President	1 Governor, 1 Vice-Governor	1 Mayor, 1 Deputy Mayor
Legislative	513 Federal Deputies (8 to 70 per state) 81 senators (3 per state)	1059 State Deputies in All Assemblies (30-49 deputies per state)	51924 municipal councillors (7-53 per municipality)

Note: This table shows the institutional framework of the Brazilian state. Source: : O. Dabéne, F. Louault; Atlas du Brésil and Ramírez (2021).

Voting in Brazil is compulsory for every citizen, born or naturalised (so the newly arrived Haitian and Venezuelan immigrants do not qualify), literate and between ages 18 to 70. If a Brazilian citizen fails to vote, then they are subject to a small fee (around US$1.50) or else they have to justify their absence to an electoral judge. Absentee citizens may also be subject to a series of penalties such as in ability to renew a passport or identity card or get a government-backed loan, among others (Freire and Turgeon, 2020). Voting is optional for people aged below 18 and 70+ as well as people who are considered illiterate.

Since my research question aims to study the impact of immigrants on natives' voting behaviour due to the location of immigrants in the same municipality, my analysis does not include the votes cast by Brazilians living abroad (2,565).[13] The reason is because I do not have information as to when the voter left their municipality of origin and so I cannot know whether or not they have been exposed to Haitian immigration.

The 2018 elections were different from the elections in earlier years. Prior to the election, Brazil, was facing a severe recession with poverty and unemployment on the rise, and a corruption scandal "Operation Car Wash" that tarnished the reputation of politicians mainly in the Workers' Party. Throughout this period, Brazil was facing a serious security crisis, with murder rates at an all-time high. The country's largest and most prominent gang, Primeiro Comando da Capital (PCC), was expanding its battles against its rivals over major drug routes.

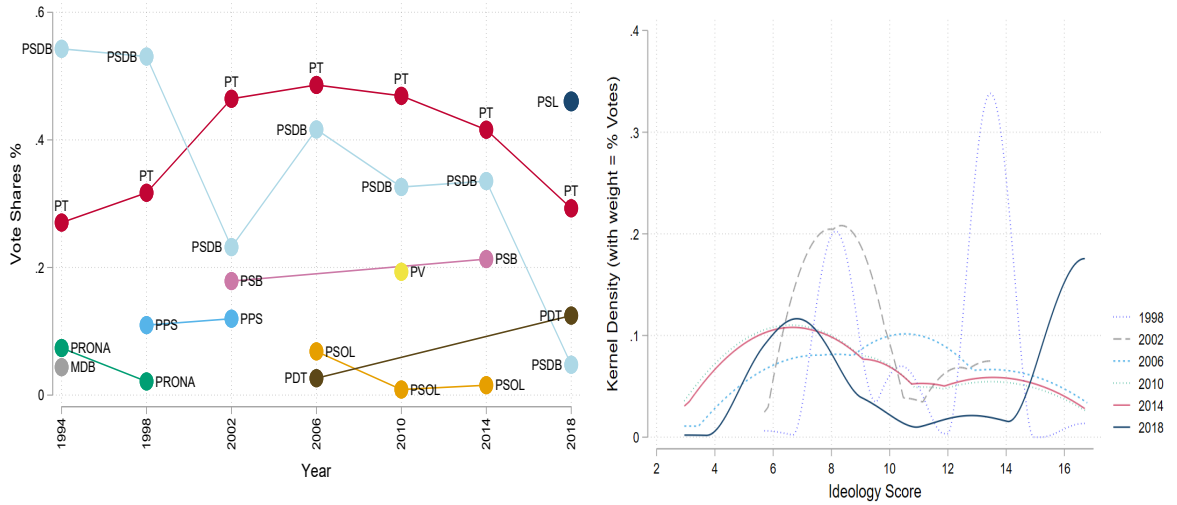
[13]The number of Brazilians who voted from abroad is small in magnitude. This can be due to the fact that for Brazilians residing abroad to vote they should have requested a transfer of their electoral domicile to the Foreign Electoral Zone (ZZ). Also voting in a foreign country can impose transaction costs that a resident in the country doesn't face.

The victory of Jair Bolsonaro in 2018 marked an ideological shift in the Brazilian politics after years of dominance by the left-wing "Workers Party" (Partido dos Trabalhadores, PT). In figure III.3(a), I plot the vote shares received by each of the top four presidential parties per year. As shown in the figure, the left-wing PT party dominated in the years after 2002. In 2018, the PSL party, with the far-right Jair Bolsonaro as party leader rose to power. In figure III.3(b), I plot the kernel density of parties across the ideology scale weighted by the number of votes given for each party. In the 2018 Presidential elections, the ideology scale for parties shifted to the right in comparison with previous years. The far right also received much higher votes than other right wing parties in earlier years. In figure III.3(a), I plot the vote shares for each election year since 1994 to 2018 for all parties that received more than 2% in the first round popular vote shares. The figure shows the dominance of the left-wing Workers Party (PT) over the centrist parties up until 2014, after which the party loses to the far-right PSL. The shift in vote shares from 2014 to 2018 towards the right-wing party is apparent in the maps I present in figure III.4.

Figure III.3: Vote Shares and Ideology of Presidential Parties Across Time

(a) Vote Shares (%) by Main Parties

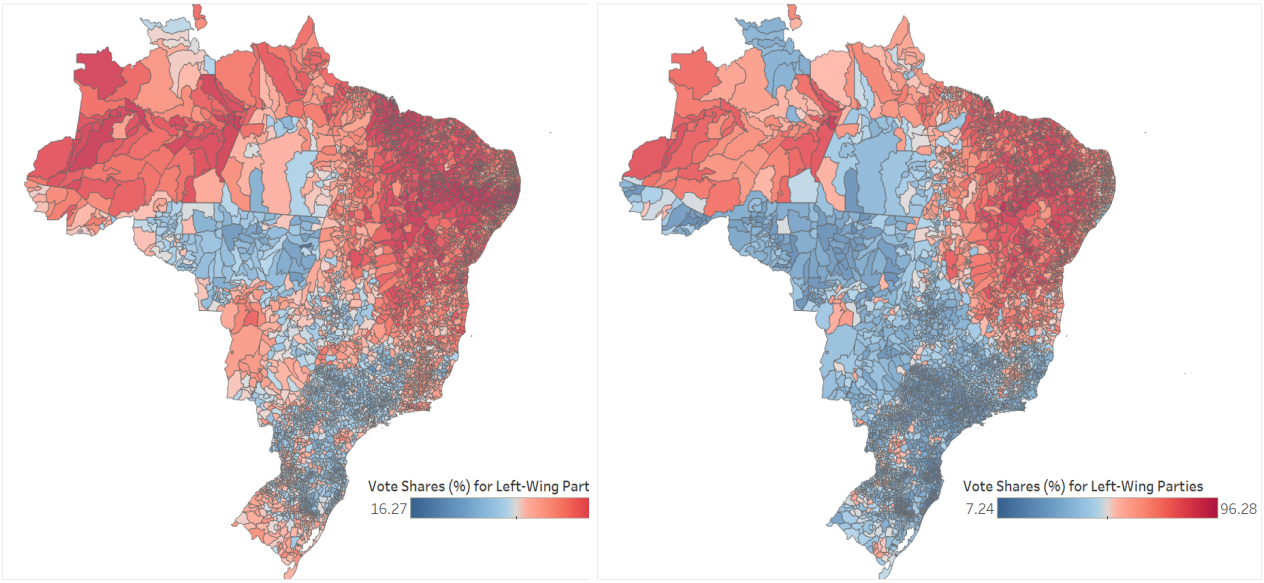
(b) Ideology of Political Parties running for Presidency Elections



Note: Figure III.3(a) shows the vote shares (%) of the parties that ran in the presidential races since 1994. The figure only depicts the parties that received at least 2% of the votes. Vote share is defined as the votes given to a certain candidate as a percent of all votes in the presidential elections. Figure III.3(b) shows the Kernel density of the ideology scores for the Brazilian Presidential elections weighted by the number of votes given. I overlay the kernel density of the ideology scores for each available year from 1998-2018, weighted by the vote shares for each party. Source: Baker and Greene (2019).

Figure III.4: Votes for Left Wing Parties

(a) 2014 (b) 2018



Note: Author's analysis of the elections data by municipality. The depicted voting outcome is equal to the ratio of votes going to the left-wing party as a proportion to total votes in the presidential elections, per municipality. Source: Tribunal Superior Eleitoral.

Jair Bolsonaro aligned himself with the PSL party, and branded himself as an anti-establishment candidate. He had been a federal congress member since 1990, with a history as an army paratrooper. Bolsonaro centred his campaign on issues of security and unapologetically made statements that were characterised by others as racists, sexist, homophobic and xenophobic. He also defended vociferously the brutal military dictatorship that ruled Brazil from the 1960s to the 1980s. To understand better the position of the main presidential candidates on different issues, I use data from Wiesehomeier, Singer and Ruth-Lovell (2021). The data, presents for each political topic like Immigration, Transfers, Privatisation, etc., a score that represents a candidate's position on this topic. Each topic has a different interpretation of the scale given to the answer, but in general, higher values of the score represent more conservative positions.[14] For, example, higher values on privatisation means that the candidate tends to oppose all state ownership of business and industry, and higher values on Taxes v. Spending means that the candidate tends to promote cutting social services to cut taxes.

[14]I also compare Jair Bolsonaro's views versus Michel Temer, his predecessor. The two candidates differ on many positions, with Jair Bolsonaro holding much more conservative views, especially as it pertains to immigration. This indicates that Jair Bolsonaro favours tough policy against immigration, much more than his predecessor.

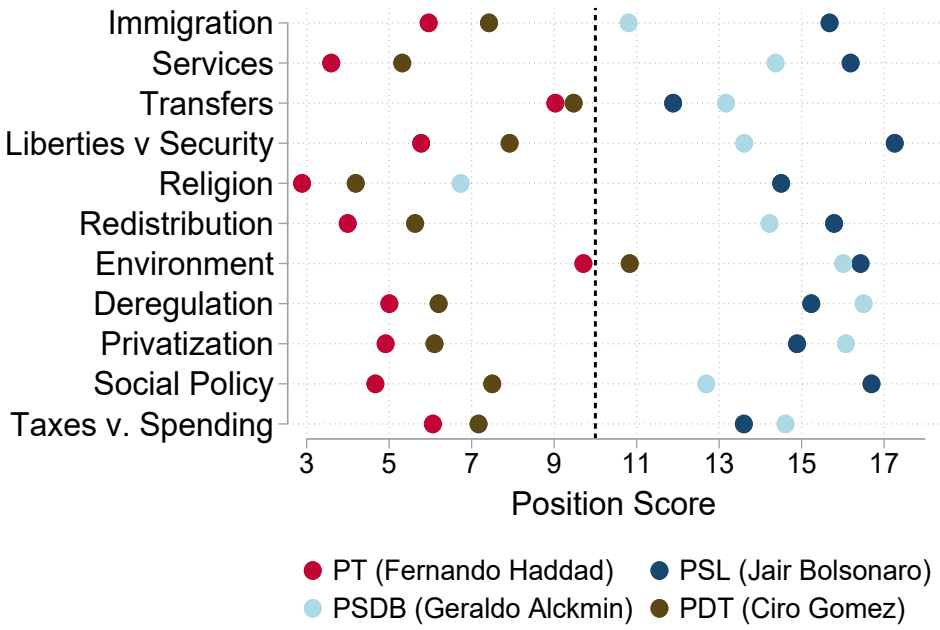
I compare the ideological positions of Jair Bolsonaro with his main competitors in the 2018 presidential elections: Fernando Haddad (left-wing party), Geraldo Alckmin (right-wing party), and Ciro Gomes (centrist party). I present the results in figure III.5. These candidates represent the top four presidential candidates in the first round of the election. Fernando Haddad ran with the "Partido dos Trabalhadores" (PT) party after being a mayor of São Paulo. This party is considered a left-wing party and has historically been one of the most important in Latin America (Van Dyck, 2014). Haddad who is of Lebanese origin, was successor to the famous and charismatic Luiz Inácio Lula da Silva after the latter was barred from running to elections by the Superior Electoral Court in response to corruption charges.[15] Jair Bolsonaro, who ran with the PSL is characterised as holding ideologically far-right views. Prior to that, Bolsonaro was a Federal Deputy for Rio de Janeiro from 1991-2018. Geraldo Alckmin ran with the party "Partido da Social Democracia Brazileira" (PSDB) after serving as the Governor of São Paulo. The PSDB party was characterised ideologically as a centrist party leaning to a right-wing ideology (Segrillo, 2004). And Ciro Gomes who ran with the party "Partido Democrático Trabalhista" (PDT) was considered to be ideologically a left-wing party. The positions across the different topics is clearly different between the candidates. Fernando Haddad, PT-workers party, aligned himself with left-leaning positions on the majority of topics, and Gomes leaned more towards the centre on the different topics. Haddad had lower values than the other two candidates on "Religion", indicating stronger support for secular politics.[16] The key topic of “Immigration” showed Haddad supporting more immigrant friendly policies towards immigration, but both Geraldo Alckmin and more so, Jair Bolsonaro favoured tougher immigration policies. Interestingly, when the topic is related to economic and redistribution policies, Geraldo Alckmin showed even more right-wing and conservative policies than Jair Bolsonaro. For instance on the “Transfers” topic, while Haddad was of the opinion that transfers should guarantee benefits to all, Geraldo Alckmin supported that transfers should be better targeted to those who need them most and based on a discretionary distribution of such benefits. This latter position is stronger than that of Bolsonaro’s. Similarly for the topic on “Taxes v. Spending”, Alckmin

[15]Bolsa Família is a nationwide conditional cash transfer program in Brazil that was introduced in 2003.

[16]The original scale on this topic “Religion” by Wiesehomeier, et al. (2021) is reversed where higher values indicate stronger support for secular politics while lower values indicate support for religious principles as a guide for politics. I subtract the highest value on scores from the position scores on religion in order to make the interpretation of the positions on this topic consistent with the other themes.

promoted cutting public services to cut taxes, more than Bolsonaro did. Consistent with Alckmin's image as a business friendly person, he opposed all forms of state ownership of businesses and called for market deregulation, more than Bolsonaro did.

Figure III.5: Ideological Positions of the Top Four Presidential Candidates



Notes: The figure shows the position of each candidate on a set of topics. The higher the position score, the more conservative the position is on each of the topics. The graph compares the ideological positions of the four top Presidential candidates in 2018: Jair Bolsonaro, Fernando Haddad, Ciro Gomez and Geraldo Alckmin. Source: Wiesehomeier, Singer and Ruth-Lovell (2021).

3.3 "Haitian Invasion": The Media and Political Narratives

The narrative surrounding Haitian immigration evolved as the number of the Haitian immigration increased over the years. In this section, I describe briefly the how the Haitian immigration was viewed in the Brazilian media as a way of shedding light on public attitudes towards. In the beginning of the Haitian immigration, around 2011 to 2012, the actors involved in the welcoming of Haitians at the border were demanding greater responsibility on behalf of the government to integrate Haitians into the Brazilian society, and to go beyond the mere practice of offering visas to providing humanitarian aid to support the immigrants (Farias, 2011). Another set of criticism was pointing towards the need to *control* the entry of Haitians. By 2012, there a was a shift in the narrative from one that sees Haitians as *escaping* into Brazil to another

that sees it as an *invasion* into Brazil (Cogo and Silva, 2016). Media reports during this time started lending support to the latter narrative through the usage of terms such as "refugee overcrowding" or the "dramatic situation", particularly in the State of Acre. This was accompanied by public demand to support the necessity of Haitians receiving basic living conditions, to which the government responded by transferring many Haitians to other municipalities within the country (Carvalho, 2012a; Lima, 2012). The narratives of "illegality" continued in Brazil, with media pointing out how the neighbouring countries like Ecuador, Colombia, Peru and Bolivia, did not allow the entry of Haitians without a visa, unlike Brazil (Carvalho, 2012b). Furthermore, the failure of the federal government's policy regarding the visa quota, to curb the fast influx of the Haitian immigration, has also intensified the narrative lending support to the "illegality" of the immigration (Frayssinet, 2012).

The situation reached a climax when the state of Acre, triggered by massive floods, closed the Haitian camps and sent to São Paulo around 800 Haitians by bus. This created a political crisis with wide media visibility (Estarque, 2014). At the centre of this political debate was the governor of Acre, Tião Vianna (left-wing PT) and Geraldo Alckmin (centrist party, PSDB). The latter accused the state of Acre of irresponsibility of sending the Haitian immigrants without a warning, while the former framed the "elite" of São Paulo as "racist" and accused them of "sanitisation". This intensified the narrative urging the re-establishment of security on the borders while exposing Brazil's geopolitical intervention in Haiti (The MINUTSAH). This is not to say that this was the only narrative. In fact counter-discourses criticised the criminalising terminology, and questioned the criteria for wanted versus unwanted immigrants, from which white and European immigrants have not been considered a problem, as opposed to what was observed in relation to the Haitians (Belchoir, 2014 as cited in Cogo and Silva, 2016).

I turn my attention next to describe the data used in this paper, the empirical and identification strategies, before presenting the results and a discussion of the channels of impact.

4 Data

I combine different data sources to conduct my analysis. I use detailed and new data on immigration into Brazil. I combine that with electoral outcomes data at the municipal level from the *Tribunal Superior Eleitoral*. Further, I use Census data from 2000 and

2010, as well as data from the Brazilian Institute of Geography and Statistics (IBGE) to create a rich set of baseline control variables at the municipal level. I provide an explanation on this data in what follows.

The data on immigration relies on databases released by the Brazil-based Observatory of International Migration (OBMigra).[17] This data is relatively new and updated regularly, and to the best of my knowledge, only a small number of published studies have made use of (Galvao and Uebel (2019); Tonhati and De Macedo, (2021); reports by OBMigra), especially to answer causally specific research questions.

I rely on two main databases. First, I use the National Migration Registry System (SISMIGRA). SISMIGRA is an administrative record of the Federal Police, and its contents refer to immigrants who have applied for registration for the issuance of the RNM (National Migration Registry).[18] The dataset makes it possible to establish the profile of the immigrant, including sex, country of birth, residence, profession, year of entry and year of registration, among others, at the municipal level. I complement this database with another that reports data on asylum requests reported by the International Traffic System Alerts and Restrictions Module (STI-MAR), under the responsibility of the Federal Police. The SISMIGRA dataset only records immigrants that applied for the National Migration Registry, and so it doesn't capture all potential immigrants, especially illegal immigrants, who are present on the Brazilian territory. This is a limitation of the dataset. To ameliorate this concern, I combine this dataset with the Asylum seekers dataset (STI-MAR).[19]

An immigrant is defined as someone who has a recorded place of birth outside Brazil and is not a naturalised Brazilian. The immigrant holds a visa status that is either permanent, temporary, resident, asylum or provisional.

[17]The data used in this paper was last accessed on June, 2022 from the official portal of the Ministry of Labour and Employment. The Observatory of International Migration (Observatório das migrações internacionais , OBMigra) is an institute for the purpose of studying international migratory patterns in Brazil, in cooperation with the Ministry of Labour and Employment (Ministério do Trabalho e Emprego , MTE), the National Council for Immigration (Conselho Nacional de Imigração , CNIg), and the Department of Latin-American Studies (Departamento de Estudos Latino-Americanos , ELA) of the Social Sciences Institute (Instituto de Ciências Sociais , ICS) at the Universidade de Brasília (UnB) (UnB n.d.)

[18]The National Migration Registry represents the immigrant's identification document in the country and is issued by the Federal Police.

[19]The other dataset available for the purposes of such a study is the dataset extracted from the Federal Police that records the entry and exit of all immigrants, including those with a tourist visa and Brazilians themselves. However, this data lacks a lot of information on the municipality of residence of immigrants, which makes it unsuitable to conduct this analysis. Future work will incorporate census data and matched employee-employer datasets to get a fuller picture on the universe of immigration into Brazil.

The granting of humanitarian visas to asylum seeking Haitian immigrants took a gradual approach with the evolution of the laws that eventually removed the quota on humanitarian visas granted (see section 3). I can observe the gradual roll out of this process in my sample. Asylum requests by Haitians with no registered place of residence in Brazil peaked in the years 2013 to 2015 while there is a spike in the number of Haitian immigrants receiving permanent visas (41,449) in 2016 compared to 13,804 and 10,083 in the previous two years with missing recorded residence. As per the stipulated law, I assume that those asylum seekers received humanitarian permanent or temporary visa status later. Therefore, and not to double count the number of Haitian migrants, I drop those Asylum seekers with missing recorded municipality from the sample and leave those who received a permanent visa, all of whom have a recorded municipality as their residence, or Asylum seekers with a recorded municipality of residence or entry.[20] Further, these Haitians with no recorded municipality of residence also have no recorded year of entry, however the majority of whom have an "inactive" visa status (90.3%). This treatment of immigrants with an inactive status is consistent with what has been done in the official annual report of ObMigra (2020),p:266. In section III.C in the Appendix, I discuss further data treatment and robustness checks. In particular, I experiment with using only the SISMIGRA dataset without the STI-MAR data or using only Haitians with visas other than Asylum. I find that my methodology and results hold in these cases.

5 Empirical Strategy

In this section, I present the empirical strategy used to study the impact of the Haitian immigration on the 2018 electoral outcomes in Brazil.

For the analysis on electoral outcomes, I focus on the 2018 election. I study the impacts at the municipality level and I estimate the following main equation:

$$E_m^j = \alpha + \beta_1 Mig_m + \beta_2 X_{m,t_o} + \theta_r + \epsilon_m, \tag{5.1}$$

[20]The majority of those Asylum seekers had an active status of an Asylum request. For migrants that report a municipality of entry but not municipality of residence, I assign to them the municipality of entry. This is a limitation as immigrants could have moved from this location to another unknown location. In order to reduce the impact of this on the analysis I perform robustness checks where I drop from the sample prominent municipality entry points for Haitians and for all immigrants (for the latter, all those above 1,000 immigrants received from 1920-2019). I find that the results are robust to the exclusion of such municipalities.

where E_m^j refers to either the turnout ratio or the vote share as a proportion of total presidential votes received by party j in municipality m. Those outcome variable are calculated as such:[21]

$$\text{Turnout Ratio} = \left(\frac{votes_{casted}}{Voters_{registered}} \right)_{2018}. \tag{5.2}$$

or

$$vote_m^j = \left(\frac{Votes_m^j}{\sum_{j=1}^{J} Votes_m^j} \right)_{2018}, \tag{5.3}$$

and, Mig_m represents the ratio of migrants in every municipality in 2018. In particular,

$$Mig_m = 100 \times \frac{\sum_{2010}^{2018} M_m}{P_{m,2010}}. \tag{5.4}$$

θ_r are mesoregion fixed effects.[22] X_{k,t_o} are lagged municipality characteristics to avoid post treatment bias include a rich set of municipality economic, social and welfare characteristics from Census data 2010, IBGE SIDRA database as well as education and demographic characteristics from TSE electoral data. I also control for the ratio of homicides to the population, drawing on municipality data from Datasus. Finally, I use data on the Bolsa Familia cash transfer program, from MI Social.[23].[24]

β_1 is estimated from comparing the fraction of immigrants to the population across municipalities within the same mesoregion in 2018. Since municipality population could in itself be influenced by immigration, historic and concurrent, the number of immigrants is scaled by the municipality population.

5.1 Identification Strategy

The estimation of equation (5.1) with ordinary least squares (OLS) will create biased estimates because there are unobservable local characteristics affecting natives' voting preferences and which could be correlated with the chosen locations of Haitian im-

[21]The treatment of the variables is analogous to the most recent literature on the impact of immigration on political outcomes (Tabellini, 2020; Rozo and Vargas, 2021; Mayda, Peri and Steingress; 2022).

[22]A mesoregion is subdivision of the Brazilian states and groups together multiple proximate municipalities with common characteristics. They are 137 in total.

[23]I use available municipality characteristics in the year 2000.

[24]The results in this paper are robust to the inclusion of different sets of variation of these controls.

migrants. One can expect that immigrants are attracted to the municipalities that promise more job opportunities, more public expenditure, more favourable political atmosphere, cheaper housing prices or better amenities. Immigrants could locate in municipalities with specific economic or demographic changes that attract immigrants and also impact local votes towards a certain party. The unexpected timing of the immigration wave that I study does not necessarily guarantee the exogeneity of the distribution of the migrants. A balance test shown in table III.30 in the appendix across electoral outcomes in 2010 for municipalities that received either Haitian immigrants does indeed show that despite the exogenous push factor of the migration episode, i.e. the earthquake, immigrants did not locate randomly. The demographic profile of Haitians also suggests that the immigrants are single men in working age and who can also be considered economic immigrants. Haitian immigration was dispersed across Brazilian municipalities, with São Paulo absorbing the majority of Haitians coming into the country. To account for this endogeneity, I rely on an instrumental variable strategy to identify the impact of the Haitian immigration. I elaborate on this in what follows.

5.1.1 Using a Shift-Share Instrumental Variable Strategy

The location of migrants across municipalities of Brazil is not random. If the unobserved factors that attract migrants to a given location, also relate to voting outcomes and electoral behaviour, then the estimates could be biased. The most popular way to address this problem in the migration literature is to use a shift-share type instrument. In this case, the shares are constructed using information on the historical location of migrants from the same country of origin and the shifts refer to the new migrants' inflows from these countries (Jaeger et al., 2018). The necessary assumption is that the factors that determined the historical location of migrants do not affect current dynamics pertaining to the outcome variable of interest.

In my context, migration to Brazil has occurred in waves. As noted above, there was a very small population of Haitians before the mid-2000s. With this in mind and to exploit more of the spatial variation of historic immigrant location, I use a modified version of the shift-share instrument which relies on linguistic proximity between incoming Haitian immigrants and historic migration language groups. The idea is that having more individuals with a common or similar language in a location can affect the location decisions of new migrants. For instance, most Haitians speak a

French-based Creole language (i.e. Haitian Creole) and could be more likely to settle in locations with a larger number of French speakers. In that sense, I replace country of origin as the common factor across migrants with language similarity based on the language spoken by the majority of the population in the country of origin. Through this method, described below, I create a Haitian unit equivalent using a measure of linguistic proximity between past migration settlements and new immigration waves.[25]

I create a Haitian unit equivalent from existing migrants using a measure of language proximity:

$$L_m = \sum_{f=1}^{F} \lambda_{f,l} (\frac{Mig_{f,m}}{Mig_f})_{t=0} \tag{5.5}$$

Where $l \in CRE$ and $\lambda_{f,l}$ is the language proximity between language Creole and language group f (e.g. French), $Mig_{f,m}$ is the number of speakers of language f (e.g. French) in municipality m at time 0 ($t_0 = 2000$).

In order to get the instrument $Z_{m,2018}$ for each municipality:

$$Z_m = \frac{(L_m * \sum_{2010}^{2018} Mig_{HTI,t}^{-m})}{\widehat{P_{m,2010}}} \tag{5.6}$$

where $P_{m,2010}$ is predicted population in municipality m in year 2010; L_m is the share of immigrants in 2000 and is calculated as per equation (5.5); and $Mig_{HTI,t}^{-m}$ is the number of Haitians that entered Brazil from 2010 to 2018, net of those that settled

[25]Note that this approach preserves the role of the historic settlements based on country of origin but also allows other countries through their proximity to Haitian Creole to influence the relevance of the IV in the first stage. I chose to work with this approach in this paper because it allows me to use more of the underlying variation in the past settlements of immigrants (instead of 48, I can use 243). Also, upon a comparison of this language proximity IV and the IV relying on earlier Haitian settlements, I find that for the full sample (only excluding main municipalities of entry of Venezuelans) the language proximity IV gives me more power in the first stage. The standard errors in the first stage of the language proximity IV are a bit smaller in the full sample than the traditional IV. Section III.B in the appendix shows these results and provides more results. Using more of the historic immigration settlements will also allow me to study different heterogeneous effects.

in municipality m.[26] [27] [28] This specification is broadly similar to the ones used in classical papers such as Card (2001) and Altonji and Card (1991). That said, I present in the Appendix III.B a brief discussion on the limitations regarding multiplicative construction of this instrumental variable.

Language is important in the formation of identity and culture of a group (Lazear, 1999). The centrality of language to economics (such as trade, migration, consumer choice etc.), sociology and political science can hardly be ignored (Ginsburgh and Weber, 2020). Bredtmann et al. (2020) study the role of networks and linguistic distance on the locations choices of migrants coming to the EU. The authors find a positive migrant networks effect and a negative impact of linguistic distance; these factors are substitutes in the location choice: networks become more important when the linguistic distance is higher, and the negative effect of linguistic distance is lessened when the network size increases. However, Bredtmann et al. (2020) study the linguistic distance that relates the language spoken in the country of origin with the host country, rather than language of earlier immigrant settlements. The role of language proximity can be relatively more important in the context of forced immigration (Bredtmann et al., 2020). Unlike forced immigrants who often have to learn and adapt on the spot to their new host country, voluntary immigrants are to some degree able to choose their destination and to apply themselves to the learning of a new language prior to leaving (Ginsburgh and Weber, 2020).[29]

In order to measure linguistic proximity, I employ the Automated Similarity Judgment Program (ASJP) that aims to establish a lexical distance between languages on

[26]Predicted municipality population, following Mayda et al. (2018) and Tabellini (2020), is calculated as such:

$$\widehat{P_{m,2010}} = Pop_{m,2000} \times \frac{Pop_{t,2010}^{-m}}{Pop_{t,2000}^{-m}} \tag{5.7}$$

.

[27]My first stage and second stage results are robust to using predicted or to using actual levels of population in 2010 per municipality. 2010 represents the year prior to the influx of immigrants.

[28]The leave one out strategy has been used in multiple studies for this instrument (such as Buchardi et al. 2019 ; Tabellini 2020). The motivation for using a leave one out strategy is to remove the mechanical covariance arising from spatial correlation of the migration shock from using own observation information. This problem is analogous to that of two-stage least squares with many instruments (Bound et al. 1995), and the solution is similar to the jackknife instrumental variable estimate approach of Angrist et al. (1999). For an elaboration on the leave one out strategy in the context of an instrument with multiple shocks, I refer readers to Borusyak et al. 2022. As robustness check, I present in table III.21 the first stage estimates where I subtract from the total flows, the total of Haitian immigrants who settled in that microregion, as opposed to the municipality.

[29]In Haiti, English is spoken by an insignificant minority of the population, which are mainly young, foreign trained and business oriented communities. Spanish is also a minority language spoken mainly due to the interactions with the bordering Dominican Republic (World Atlas, 2019).

the basis of standardised wordlists[30]. The recent list from June 2015 covers nearly 6,895 languages and dialects. The algorithm makes use of the Levenshtein (1966) distance. The Levenshtein distance is the minimum number of insertions, deletions, or substitutions of a single character to transform a word in one language into a word in another language (Serva and Petroni, 2008). Further, these languages are annotated based on the classifications used by the World Atlas Language Structures Online (Muller et al. 2010), which also assigns the languages to countries. I standardised this measure and subtracted it from 1 in order to create a measure of linguistic proximity between two languages. Then, I use the CIA The World Factbook and the World Atlas Languages Structures to assign, for each country, the language spoken by the majority of the population.[31] I then use the linguistic proximity vector to construct the shift share instrument.[32]

I plot, in figure III.6, the linguistic proximity of different language groups to Haitian Creole.[33] I find that the majority of languages have a proximity index to Creole that varies between 0 and 0.2, while a handful of countries, including French, have a higher index between 0.4 and 0.6. It is possible that this latter group of countries would be disproportionately driving the underlying variation in the instrumental variable strategy. However, the proximity index is multiplied by the share of these language groups that were actually present in Brazil prior to 2000, and hence each language group will be weighted differently as a result. Furthermore, I present a formal analysis of the underlying weights of each language share as per Goldsmith-Pinkham (2020) in the Appendix and in section 7, I discuss the importance of testing different language proximity indices.

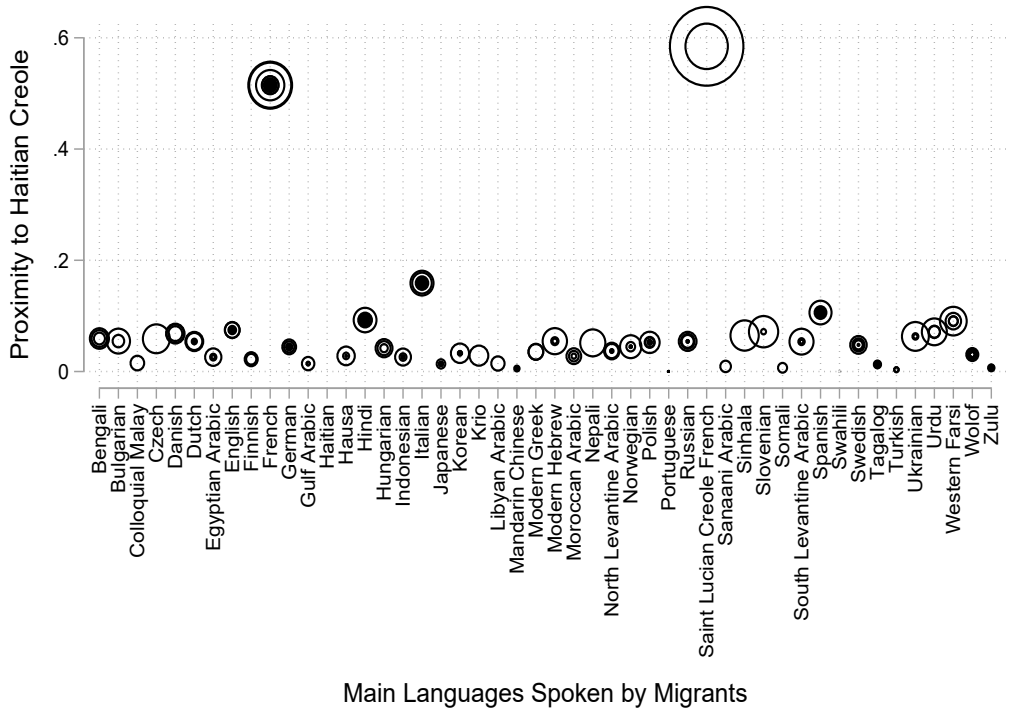
[30]For a description of The ASJP Database, I refer readers to Wichmann et al. 2020.

[31]I present in the conclusion a discussion of how the matching between the language spoken and the country of the immigrant can be further improved.

[32]Although, this doesn't affect the first stage results and the power of the IV for the full sample, nor the results, I exclude from the construction of the shift share, the top extreme outlier municipalities where Z_m is greater than the top 1 percentile. Those are only three municipalities Bonfim, Macae and Poco dos Antos, out of more than 5558 municipalities used in the analysis. This is driven by either a very large historic share (25% as is the case for Macae) or a small population size (case of Bonfim and Poco Dos Antos). Bonfim also represents a main municipality of entry into Brazil for immigrants. I also exclude the two main entry municipalities and residence of Venezuelans (Pacaraima and Boa Vista).

[33]The main language spoken in Brazil is Portuguese, therefore it makes less sense that this particular language group will be determining the location of incoming Haitian immigrants. Therefore I replace this proximity index between Portuguese and Haitian Creole by zero, and hence not allowing Portuguese immigrants to affect the locations of future Haitian immigrants.

Figure III.6: Proximity to Haitian Creole $\lambda_{f,creole}$ by Historical Immigrant Spoken Languages (before 2000)



Notes: The figures show proximity to Haitian Creole by each country and language group. The proximity to Haitian Creole is normalised measure with higher values indicating more proximity. Haitian Creole proximity to itself is 1 and is omitted from the graph. The language chosen for each country is the language spoken by the majority of the population in the country.

5.1.2 Identification Assumptions, Relevance and Validity of the IV

My identification strategy relies on the language-proximity of earlier migrant settlements to new migration waves. The validity is based on the assumption that the 2000 migration shares varied across municipalities because of factors related to the persistent agglomeration of migrants in some municipalities, but, after controlling for a rich set of municipality characteristics, were unrelated to shocks affecting municipality level voting outcomes. The IV as per equation (5.6) exploits one major source of variation: A cross sectional variation in the share of immigrants from each language group living in different municipalities in Brazil, L_m. I portray this variation graphically in figure III.11 in the appendix, where I show the distribution of the Haitians and the calculated IV across municipalities, and the correlation between the two. I also plot in figure III.12, the variation within mesoregion as this is important because my empirical strategy exploits only within-mesoregion variation; if the latter doesn't exist my instrument will not have any power. In the same figure (III.12), I show the variation in a different way, where I plot the percent of municipalities with at least one Haitian immigrant per mesoregion.

I present the first stage results in table III.3. The first stage is significant and robust to multiple specifications. The IV first stage remains strong for a population-weighted regression as well as the unweighted regression. The first stage remains strong either with the full sample or a subset of the sample including only those municipalities that received at least one immigrant.

However, there are multiple threats to the validity of the instrument. First, since my identification is driven by the exogeneity assumptions surrounding the shares component of the IV, i.e. L_m in (5.5), I address the recent critiques on shift-share instruments by Goldsmith-Pinkham et al. (2020) and provide a detailed analysis in the Appendix. I find that the top three language groups in terms of their contribution to the variation and power of the instrument are Haitian Creole, Spanish and English. I perform further robustness checks on these language shares to check if they are correlated with confounding baseline municipality characteristics that could also predict political outcomes. I portray the results in table III.16. I find that for the majority of the baseline characteristics, there is no statistically significant relationship with any of those shares. I do find a statistically significant relationship between those shares and the overall IV with value added in services. The coefficient however is small and the adjusted R^2 doesn't exceed 0.04 in any of the regressions. I make sure to control for

these variables, among others, in a rich set of baseline municipality characteristics, described in table III.31. Further, my analysis is not subject to recent critiques by Jaeger et al. (2018) because I rely on one source of immigrant shock, that of the Haitians. The Haitian shock was sudden, and negligible prior to the Earthquake in 2010 that shook Port-au-Prince.

Second, the identifying assumption behind this IV is that the municipalities in 2000 that received Haitian immigrants were not on different trends concerning economic growth and political changes than those that didn't (Borusyak et al., 2018; Goldsmith-Pinkham et al., 2020). Therefore it is important to deal with potential threats to this assumption. To begin with, if the characteristics of municipalities in 2000 that were home to early immigration settlements had certain characteristics that are persistent over time and can potentially impact my outcome variables of interest as well as the location of Haitian immigrants, then the identifying assumption will be violated. In order to address this concern, I do the following. I test if historical migrant shares are correlated with pre-trends in the outcomes leading up to the major Haitian immigration shock. I present the specific estimated equation and results in the appendix (equation (III.B.15) and report the results in figure III.13 and table III.17). I find that the instrument with time dummies up until 2006 (before the onset of the immigration), had no significant impact on turnout ratio, votes for the left-wing party and votes for the centrist party. In table III.17, I show that the 2000 language shares had no significant correlation with the turnout ratio in the years leading up to the immigration. Further, in columns (4) to (5) in table III.17, I show that the turnout ratio in the years prior to the immigration is not correlated with the future location decisions of Haitian immigration in 2018. Following Tabellini (2020), I perform another set of tests to deal with the concern of confounding and persistent municipality characteristics that might impact my outcome variables and migrant location decisions in 2018. In particular, I show that the pre-trend change (from 2000 to 2010) of some of my key outcome variables is not correlated with the subsequent instrumented immigration variable. I present the results in table III.18 and III.19 in the Appendix.

Third, the IV relies on the historic concentration of immigrants from a certain language group in a given municipality. These may tend to be municipalities that have a large share of immigrants. One would worry that such immigrant stocks could influence the political arena or labour markets in such municipality. I don't believe this concern holds in this context, mainly because early settlements of migration exceeded

1% or total municipal population for only seven municipalities in 2000.[34] When I exclude those municipalities from my estimation, I find the first-stage point estimates and power do not change. Further, I control for 2000 native population and immigrant population per municipality and I find that the first stage remains strong. The results are present in table III.4

Fourth and finally, I address potential concerns surrounding the validity of the instrument and the viability of the language proximity index. I run a series of estimations presented in table III.5 to test the sensitivity of the instrument to the sample composition and the language proximity index. I test whether the first-stage results are driven by municipalities that are at the border, those with large areas and large population, with results presented in columns (1)-(3). I find that the first-stage estimation remains highly significant and the Kleibergen-Paap F-statistic is comfortably large when such municipalities are excluded. Column (4) tests the importance of the instrumental variable assumption in terms of whether new waves of migrants locate near earlier migrant settlements that share their language. As a placebo test, I attempt to predict recent waves of Haitians using a vector that measures language proximity to Japanese instead of French Creole.[35] The results are presented in column (4) where I find that this instrument gives a weak first stage, a result that is consistent with my hypothesis that language proximity matters to predict migrant locations. Weighting the language share by the proximity to the Japanese language instead of the Haitian Creole also reduces the influence of Haitian earlier settlements, which from my analysis (table III.15), are key in the weight of the Bartik IV first stage estimate. I experiment (not shown) by keeping the weighting through the proximity of Haitian Creole to all immigrants, but instead of weighting the Haitians by 1, because the proximity to Creole is 1, I weight them by 0.5, similar to other French speakers. I find that my first stage remains strong with a Kleibergen-Paap F-stat of 35.3. Columns (5)-(6) verify that the validity and power in the first stage is not dependent on the locations of migrants with permanent visas, held disproportionately by immigrants from richer countries, nor by migrants from Mercosur countries, respectively. Another potential concern is that the inflow of Haitians into certain municipalities is a proxy for aggregate labour demand growth or certain political trends that are attractive for migrants, perhaps those similar in profile to Haitians (Other Asylum seekers). Column (7) tests whether

[34]The ratio of immigrants to population up until 2000 had a mean of 0.00107 and a standard deviation of .0039.

[35]Instead of using how close is the main language of earlier settlement groups is to Haitian Creole $\lambda_{f,Creole}$, the main language used in Haiti, I use the proximity to Japanese instead $\lambda_{f,Japanese}$.

the instrument is also predicting other asylum seekers and not just Haitian migrants. However, I find that this is not the case as the first stage point estimates from column (7) are insignificant.[36]

Table III.3: First Stage Results

	(1) HTI Flows (interval)/Pop 2010	(2) HTI Flows (interval)/Pop 2010	(3) HTI Flows (interval)/Pop 2010	(4) HTI Flows (interval)/Pop 2010
IV First Stage: Haitian Migration				
IV HTI shift (2010-2018)/pop 2000 Language	21.69***	21.51***	19.99***	20.06***
	(4.127)	(4.044)	(3.744)	(3.688)
Observations	5565	5565	5558	5558
Year	2018	2018	2018	2018
Sample	All	All	All	All
Weighted Regression	Population Weighted	Nonweighted	Population Weighted	Nonweighted
Kleibergen-Paap F-stat	27.6375	28.2826	28.5050	29.5905
Cragg-Donald F Stat	1.0e+03	948.1130	872.1200	813.0726
Municipality Controls	No	No	Yes	Yes
Mesoregion F.E	No	No	Yes	Yes

Notes: The table shows the 2SLS first stage estimates. Robust standard errors in parentheses clustered at the mesoregion level. Regressions include mesoregion fixed effects and municipality controls in 2010. Results are shown for two sample: full sample of municipalities and municipalities that received at least one migrant. The predicted Haitian migration variable is adjusted by population in 2010. Data sources: IBGE 2010 Census data for control variables, Tribunal Superior Eleitoral and STI-MAR and SISMIGRA for migration data.

[36]It is worthwhile to note that I also test whether the IV is predicting locations of other migrants with different visa status (permanent, temporary and resident) and find that the first-stage is weak. I present those results in table III.20. This result is consistent with the hypothesis that the instrument is not conflating local labour demand factors

Table III.4: First Stage Results: Robustness Checks I

	(1) Keep Mun with Zero Migrants	(2) Excluding Mun with $(\frac{Mig}{pop})_{2000} > 1\%$	(3) Control 2000 Pop	(4) Control 2000 Pop and Haitian Pop	(5) Control 2000 Pop and Mig Pop
IV First Stage: Haitian Migration					
IV HTI shift (2010-2018)/pop 2000 Language	20.27***	20.71***	20.10***	18.66***	20.09***
	(3.669)	(3.854)	(3.763)	(3.764)	(3.763)
Population 2000			-2.84e-08**	-0.000000126***	-5.69e-08*
			(1.27e-08)	(2.91e-08)	(2.99e-08)
Initial Haitian Pop				0.0631***	
				(0.0185)	
Other Migrants Pop					0.000000925
					(0.000000854)
Observations	5299	5550	5558	5558	5558
Year	2018	2018	2018	2018	2018
Sample	All	All	All	All	All
Kleibergen-Paap F-stat	30.5276	28.8680	28.5366	24.5720	28.5084
Anderson-Rubin Test	0.2795	0.3729	0.2729	0.1921	0.2898
Cragg-Donald F Stat	842.6115	913.6346	878.9201	684.2226	877.4357
Municipality Controls	Yes	Yes	Yes	Yes	Yes
Mesoregion F.E	Yes	Yes	Yes	Yes	Yes

Notes: The table shows the 2SLS first stage estimates. Robust standard errors in parentheses clustered at the mesoregion level. Regressions include mesoregion fixed effects and municipality controls in 2010. Results are shown for the full sample of municipalities under different specifications. The predicted Haitian migration variable is adjusted by population in 2010. Data sources: IBGE 2010 Census data for control variables, Tribunal Superior Eleitoral and STI-MAR and SISMIGRA for migration data.

Table III.5: First Stage Results: Robustness Checks II

	(1) Excluding Border Mun	(2) Excluding Mun with Large Areas	(3) Excluding Mun with large pop	(4) Without Language transformation	(5) Without Top Permanent Visa	(6) Without Mercosur	(7) Other Asylum
IV First Stage: Haitian Migration							
IV HTI shift (2010-2018)/pop 2000 Language	18.70***	16.97***	19.85***				1.248
	(3.845)	(4.163)	(3.772)				(0.930)
IV Language - Japanese				6.059			
				(4.624)			
IV - No Mercosur					20.42***		
					(3.848)		
IV - No Top Permanent Visa						23.29***	
						(3.194)	
Observations	4976	5001	5003	5558	5558	5558	5558
Year	2018	2018	2018	2018	2018	2018	2018
Sample	All	All	All	All	All	All	All
Kleibergen-Paap F-stat	23.6517	16.6089	27.7039	1.7165	28.1442	53.1498	1.8001
Anderson-Rubin Test	0.5486	0.0244	0.4038	0.6545	0.2505	0.4567	0.4479
Cragg-Donald F Stat	1.0e+03	424.2924	782.7824	10.6792	895.9750	1.0e+03	72.8038
Municipality Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mesoregion F.E	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The table shows the 2SLS first stage estimates. Robust standard errors in parentheses clustered at the mesoregion level. Regressions include mesoregion fixed effects and municipality controls in 2010. Results are shown for the full sample of municipalities under different specifications.Data sources: IBGE 2010 Census data for control variables, Tribunal Superior Eleitoral and STI-MAR and SISMIGRA for migration data.

6 Results

In this section, I present the results of this paper. I organise the analysis along four axes. First, in section 6.1, I present the set of results related to the impact of Haitian immigration on compulsory and voluntary voter participation, defined by the turnout ratio, disaggregated by age, education and gender. Second, I present in section 6.2 the impact of the Haitian immigration on electoral vote shares. Within this section, I present the analysis for different ideological categories as well as for individual Presidential candidates.[37] Third, in section 6.3, I provide evidence on the heterogeneity of these impacts depending on certain baseline (in 2010) municipality characteristics. Fourth, in section 6.4, I attempt to shed light on the mechanisms through which the observed impacts are realised and provide a discussion synthesising the findings in section 6.5.

6.1 Impact on Voter Turnout Ratio and Composition of Voters

My analysis then proceeds to study whether the Haitian immigration affected turnout. A contribution of this analysis lies in being able to disaggregate the turnout ratio by voter characteristics such as age, gender, and education level. Therefore, the subsequent analysis sheds light on whether Haitian immigration altered the actual voter composition in the elections.

I start with a general analysis, where I study the impact of the Haitian immigration on the overall turnout ratio, as well as on turnout ratio disaggregated by gender. The overall turnout ratio is as calculated as per equation (5.2), and represents the percentage of registered voters who cast their votes.[38] I report the results in table III.6. I find that a one percentage point increase in the ratio of Haitian immigration to the population leads to a 0.97 percentage point reduction in turnout ratio. Further, the turnout for both females and males is adversely impacted due to an increase in

[37] I describe how I group the different parties into ideological categories in section III.A in the Appendix.

[38] While this is a satisfactory measure for the purposes of this analysis, it is one that doesn't account for individuals who are eligible to vote and chose not register. Because of that, such a measure can overestimate turnout. There are other measures that can be tested in the future. Those include using as a denominator the voting age population (VAP). However, the latter can be a poor proxy for the voting eligible population (VEP) because there could be a lot of ineligible voters and foreigners, that exceed the number of Brazilians living abroad. I refer readers to Evans et al. (2018) for an important discussion on the pitfalls of the mismeasurement of this variable.

the ratio of Haitian immigration, however the coefficients are not significantly different from one another.

To understand further how the voter composition has been impacted due to the Haitian immigration presence, I study the turnout ratio, disaggregated by gender and education levels. I report the results in table III.7. I use two aggregated education groups: "Up to Incomplete High School", which includes all individuals that are literate or have had formal education, but didn't complete high school; and "At least Complete High School", which includes all individuals who at the very minimum completed high school.[39] My results are consistent with the previous findings whereby an increase in Haitian immigration led to a decrease in turnout ratio, across the two genders and across education levels. Although the estimates are statistically insignificant from one another, the reduction in turnout ratio for those who are less educated (up to incomplete high school) is slightly larger than those with a higher level of education. And in the case of male voters, the lower turnout ratio is driven by those with a lower level of education.[40]

Further, I estimate the impact of the Haitian immigration on the voter turnout for disaggregated education categories across four age groups. It is worth noting that in Brazil, voting is compulsory for all people between the age 18-70 years old and who are not illiterate. Voting is hence optional for those who are illiterate, between the ages 16-18 and above the age of 70 (Tribunal Regional Eleitoral, 2021).[41] I report the results in tables III.45, III.46 and III.47 in the Appendix. I find that the reduction in turnout ratio is observed across different education levels. I find that for females, the impact on turnout ratio is driven by with a completed primary education and with the age range of 19-39 years old. Within this age group, I also note that women who are

[39]Specifically, the two categories are calculated as such. "Up to Incomplete High School" has voters who are literate, or have formal education such as uncompleted primary school, primary school, and uncompleted high school. "At least Complete High School" includes individuals who have formal education that includes either a completed high school, uncompleted university and university.

[40]I also study the impact of the Haitian immigration on temporary transfer turnout ratio and disability turnout ratio all in percent (%). The temporary transfer turnout ratio represents the ratio of voters that are outside their electoral domicile at the election day, and request to vote in special ballot boxes designated for them in municipalities with more than a 100,000 voters that voted as a percent of all registered voters. This is only permitted for the general elections of the President, Governor, Senator, Federal Deputy and State Deputy. The disability turnout ratio represents the percent of voters with a disability or reduced mobility, considering those who actually voted as a percent of all registered voters. Voters with disability or reduced mobility can request to vote in a more accessible location. I do not find any impact of Haitian immigration on temporary vote transfer, or on the participation of voters with qualifying disabilities.

[41]I study the impact of the Haitian immigration on the turnout of compulsory and voluntary voters. I find that the reduction in turnout ratio is driven mainly by compulsory voters, while I find insignificant results on the turnout of voluntary voters.

illiterate (so it is voluntary for them to vote) also reduced their turnout in response to the Haitian immigration. Further, the impact of turnout ratio is also driven by females between the ages of 40-69 who did not complete primary education and by those who completed high schools and university level education.

For males, the results are similar to females with two differences. For males within the age range of 19 to 39 years old, the impact on turnout ratio is driven by voters who have completed university and by those with an incomplete primary or high school degree. The second difference lies in the finding that an increase in Haitian immigration led to an increase in the voter turnout among the male illiterate voters aged between 40 to 59 years old, and the voters above 70 years old who hold a university degree and those with an incomplete high school degree. [42]

Table III.6: Turnout Ratio Results

OLS Estimates						
	(1)	(2)	(3)	(4)	(5)	(6)
	Turnout Ratio	Female Turnout Ratio	Male Turnout Ratio	Turnout Ratio	Female Turnout Ratio	Male Turnout Ratio
HTI Flows (interval)/Pop 2010	-0.184	-0.219	-0.146	-0.155	-0.169	-0.139
	(0.175)	(0.167)	(0.188)	(0.214)	(0.206)	(0.229)
IV 2SLS Estimates						
	(1)	(2)	(3)	(4)	(5)	(6)
	Turnout Ratio	Female Turnout Ratio	Male Turnout Ratio	Turnout Ratio	Female Turnout Ratio	Male Turnout Ratio
HTI Flows (interval)/Pop 2010	-0.970**	-0.901**	-1.050*	-0.856*	-0.852*	-0.867
	(0.485)	(0.433)	(0.561)	(0.507)	(0.484)	(0.568)
Election Round	1	1	1	2	2	2
Mean of Dep. Variable	79.18	80.04	78.30	77.23	77.71	76.75
Mesoregion F.E	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The table show the OLS and 2SLS second stage estimates for the impact of migration on turnout ratios for all voters (for compulsory and voluntary voters). The results are shown for all the sample of municipalities with available control data for the two rounds of elections in 2018. The outcome variable in columns (1) and (4) is $turnoutratio = \frac{votes_{cast}}{Voters_{registered}}$. The outcome variables in columns (2), (3), (5) and (6) represent the turnout ratio for females and males. The latter turnout ratio represents all the votes cast by females or males as a proportion of all registered males or female voters. Robust standard errors in parentheses clustered at the mesoregion level. Regressions include mesoregion fixed effects and municipality controls in 2010. Data sources: IBGE 2010 Census data for control variables, Tribunal Superior Eleitoral and STI-MAR and SISMIGRA for migration data.

[42] In addition to the other robustness checks to increase validity in the exclusion restriction, I perform tests to decrease the probability of reverse causation and the persistence of voter behaviour over time by seeing if the turnout ratio in years prior to the onset of the immigration (pre-2010) predicts the location of incoming Haitians over the *next decade*. I present the results in table III.17 in the appendix where I find that there is no significant correlation for the years prior to 2010. Further, it can be argued that unlike voting shares to a certain party, voter turnout can be less easily observed – especially by newly arriving immigrants, so it is less likely that they make their location decisions based on this factor per municipality.

Table III.7: Turnout Ratio Results by Education Level and Gender

OLS Estimates	**Total**		**Female**		**Male**	
	(1)	(2)	(3)	(4)	(5)	(6)
	At least Complete High School	Up to Incomplete High School	At least Complete High School	Up to Incomplete High School	At least Complete High School	Up to Incomplete High School
HTI Flows (interval)/Pop 2010	-0.124	-0.434**	-0.229	-0.450**	0.00144	-0.424**
	(0.163)	(0.167)	(0.149)	(0.181)	(0.187)	(0.165)
IV 2SLS Estimates	**Total**		**Female**		**Male**	
	(1)	(2)	(3)	(4)	(5)	(6)
	At least Complete High School	Up to Incomplete High School	At least Complete High School	Up to Incomplete High School	At least Complete High School	Up to Incomplete High School
HTI Flows (interval)/Pop 2010	-0.948*	-0.965*	-0.824*	-0.890*	-1.107	-1.061**
	(0.572)	(0.496)	(0.493)	(0.532)	(0.682)	(0.523)
Election Round	1	1	1	1	1	1
Mean of Dep. Variable	85.61	76.78	86.44	77.17	84.36	76.42
Mesoregion F.E	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The table show the OLS and 2SLS second stage estimates for the impact of migration on turnout ratios for all voters, dis-aggregated by gender and education levels. The first panel represents the results on turnout ratios for compulsory voters and the second panel represents results on turnout ratios for voluntary voters. The results are shown for all the sample of municipalities with available control data for the first round of elections in 2018. The outcome variable in columns (1) and (4) is $turnoutratio = \frac{votes_{cast}}{Voters_{registered}}$. The outcome variables in columns (2)-(6) represent the turnout ratio for females and males, dis-aggregated by education levels. The latter turnout ratio represents all the votes cast by females or males for each education classification as a proportion of all registered males or female voters, within that education category. Robust standard errors in parentheses clustered at the mesoregion level. Regressions include mesoregion fixed effects and municipality controls in 2010. Data sources: IBGE 2010 Census data for control variables, Tribunal Superior Eleitoral and STI-MAR and SISMIGRA for migration data.

6.2 Election Results

This section presents the results of the analysis from estimating the 2SLS equation (5.1). I first start by showing the impact of the presence of Haitians on voting outcomes for the Presidential elections. I study this aggregating the parties across ideological classifications (far-left, left, centre, right and far-right parties) and show the OLS and 2SLS results in table III.8 for the first round of elections.[43] I notice that the results from the OLS and IV 2SLS estimation are different qualitatively and quantitatively. Since Haitian immigrants locate non randomly, the bias in the OLS might be larger than the truth resulting in the estimates of opposing signs. Further, IV 2SLS is estimating a local average treatment effect for compliers and therefore this estimate may not have a direct link to the OLS ones.[44] Based on the 2SLS estimates, I find that an increase of a one percentage point in Haitian immigration increases votes for the right-wing (significant at the 5% level) by 0.89 percentage point.[45] Interestingly, there is no impact of the Haitian immigration on voting for the far-right. I find no impact of an increase in the the Haitian immigration ratio to population on second round voting in the Presidential elections.

Second, I focus my attention on the impact of Haitian immigration on voting outcomes on individual Presidential candidates. I perform this analysis on the vote shares, defined as the votes cast to a certain candidate as a ratio of total votes cast towards Presidential parties, as described in equation (5.3).[46] I portray the results in table III.9. I find that there is no impact of the Haitian immigration on vote shares for Jair Bolsonaro and Fernando Haddad. I find an increase in vote shares to Geraldo Alckmin (classified as right-wing candidate) that is significant at the 5% levels. I also

[43]The exact classification of the ideological classifications is explained in detail in section III.A in the Appendix.

[44]It could also be the case that the exogeneity assumption of the IV strategy is violated. I rule out this explanation because I perform multiple checks to test, albeit indirectly, the exogeneity assumption (Angrist et al., 1996; Gennetian et al., 2008) in section III.B in the Appendix. I also do not have a weak instrument problem (Bound et al., 1995) as evidenced by the first stage results presented in tables III.3 to III.5. The more likely explanation, if I were to take the effects on votes for the left wing party. The OLS estimate is positive indicating that immigrants potentially locate intentionally to areas where the political environment is favourable to them, so the bias is positive and potentially large.

[45]This may represent a large increase considering the mean of the vote shares of right wing parties across municipalities. However, to put this in context, this impact is in response to a one percentage point increase in the ratio of Haitian immigration to the population, which represents approximately a 42% increase from the mean (a ratio of 0.022% of Haitian immigration relative to the population) of this independent variable.

[46]The denominator ratio in vote shares is total votes cast towards Presidential parties in order to isolate the impacts on blank and invalid votes. However, I repeat the estimation of the results defining the vote share with the denominator being all voters (including those with blank and invalid votes) and the results are qualitatively and quantitatively equivalent.

find an increase in vote shares (significant at the 10% level) towards candidates in right-leaning centre parties, Henrique Meirelles for the MDB party (the minister of economy), and Joao Amoedo for the NOVO party, (a banker and advocate of less government interference).[47]

Third, the outcomes observed can be masking different voting behaviour. I check whether Haitian immigration caused a change in the share of blank or null votes. I calculate the share of blank and null votes in two ways. The first represents the ratio of blank (null) votes to total number of voters; the second approach uses the ratio of blank (null) votes to cast votes. I report the results for the first and second round elections in table III.35 in the appendix. I find that Haitian immigration impacted negatively the share of blank votes for presidential, governor, senator and state deputy elections but none of the impact is statistically significant. In the second round elections, I find that Haitian immigration increased blank votes share for presidential and governor elections, but the impact is statistically insignificant.[48]

There could be two margins impacting the effect observed on the vote shares of the presidential candidates. The first is that the lower political participation result reported in response to an increase in the ratio of Haitian immigration to the population, can be driven by an electorate supporting certain candidates not showing up for the voting booth, a *mobilisation* effect. Thus voters that actually do show up are supportive of a certain political party. The second margin could be due to voters *switching* their choices towards other candidates in response to the Haitian immigration shock. In the estimations reported in tables III.8 and III.9, the vote share has, as a denominator, total votes cast, as per equation (5.3). However, votes cast can

[47]The method of classification for all parties is described in Appendix III.A. For the PSDB party and their candidate Geraldo Alckmin, I further rely on the ideological positions across multiple themes by Wiesehomeier et al. (2021), Baker and Greene (2019) ideological scores and on the discussion presented in Lima (2012).

[48]In addition to the impacts shown in this section, I have also attempted varying the control group to which Haitian-rich municipalities are compared to. I do two different estimations: I compare Haitian-rich municipalities with municipalities that received no immigrants, and I compare municipalities that received only Haitians with those that received no immigrants. In both cases, the results of on the vote shares of the presidential candidates and turnout ratio hold qualitatively similar to what has been found in table III.9. Interestingly, when comparing municipalities that received only Haitians with those that didn't receive any immigrants, I find a significant increase in vote shares to Jair Bolsonaro as well and a decrease to Fernando Haddad. Results on turnout ratio for the latter exercise remain qualitatively similar, albeit insignificant. But due to the limited number of Haitian municipalities in this exercise, it is not my preferred specification for this paper, and leave it as a robustness check. The result on political participation in this robustness check lends support to the argument that a potential channel of impact on political participation is through a change in the heterogeneity of the voter's community. However, to provide more conclusive evidence of this channel, more investigation is needed where the ethnic heterogeneity of the Brazilian citizens and immigrants from different countries is taken into consideration.

Table III.8: Disaggregated Party Classifications: IV 2SLS Estimates for First Round Elections

OLS Estimates					
	(1)	(2)	(3)	(4)	(5)
	Far-left Parties	Leftwing Parties	Centre Parties	Rightwing Parties	Far-right Parties
HTI Flows (interval)/Pop 2010	-0.00745	0.626	-0.0321	0.301	-0.888***
	(0.00861)	(0.455)	(0.0510)	(0.201)	(0.336)
IV 2SLS Estimates					
	(1)	(2)	(3)	(4)	(5)
	Far-left Parties	Leftwing Parties	Centre Parties	Rightwing Parties	Far-right Parties
HTI Flows (interval)/Pop 2010	-0.000798	-0.839	0.0878	0.890***	-0.137
	(0.0169)	(1.094)	(0.0895)	(0.301)	(0.959)
Observations	5558	5558	5558	5558	5558
Mean of Dep. Variable	0.343	50.81	1.704	7.695	39.45
Mesoregion F.E	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes

Notes: The table shows the OLS and 2SLS second stage estimates for ideological subcategories (far-left, leftwing, centre, rightwing and far-right). Robust standard errors in parentheses clustered at the mesoregion level. Regressions include mesoregion fixed effects and municipality controls in 2010. Results are shown for the full sample of municipalities in the first round of the elections. Data sources: IBGE 2010 Census data for control variables, Tribunal Superior Eleitoral and STI-MAR and SISMIGRA for migration data.

Table III.9: Vote Shares to Presidential Candidates: IV 2SLS Estimates for First Round Elections

First Round							
	(1)	(2)	(3)	(4)	(5)	(6)	
	Jair Bolsonaro (PSL)	Fernando Haddad (PT)	Marina Silva (REDE)	Geraldo Alckmin (PSDB)	Ciro Gomes (PDT)	Jose Maria Eymael (DC)	
HTI Flows (interval)/Pop 2010	-0.0356	-0.415	-0.0329	0.555**	-0.423	-0.00191	
	(0.975)	(0.890)	(0.0270)	(0.248)	(0.294)	(0.00576)	
Observations	5558	5558	5558	5558	5558	5558	
Election Round	1	1	1	1	1	1	
Mesoregion F.E	Yes	Yes	Yes	Yes	Yes	Yes	
Controls	Yes	Yes	Yes	Yes	Yes	Yes	
First Round							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Henrique Meirelles (MDB)	Joao Amoedo (NOVO)	Cabo Daciolo (PATRI)	Alvaro Dias (PODE)	Joao Vicente Goulart (PPL)	Guilherme Boulos (PSOL)	Vera Lucia (PSTU)
HTI Flows (interval)/Pop 2010	0.121*	0.231*	-0.102	0.106	-0.00121	0.000830	-0.00163
	(0.0731)	(0.136)	(0.0658)	(0.0718)	(0.00307)	(0.0172)	(0.00386)
Observations	5558	5558	5558	5558	5558	5558	5558
Election Round	1	1	1	1	1	1	1
Mesoregion F.E	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The table shows the 2SLS second stage estimates on presidential vote shares. In this specification, presidential vote shares are defined as votes cast to certain candidate as a ratio of the number of registered voters per municipality. Robust standard errors in parentheses clustered at the mesoregion level. Regressions include mesoregion fixed effects and municipality controls in 2010. Results are shown for the full sample of municipalities. Data sources: IBGE 2010 Census data for control variables, Tribunal Superior Eleitoral and STI-MAR and SISMIGRA for migration data.

itself be subject to differences in electoral participation (as studied in section 6.1). In order to understand better how these two effects could be impacting the results on vote shares, I first study the impact on votes cast as ratio of *registered* voters, and second, I attempt a decomposition of the two effects on the vote shares.

First, to isolate the effect of voter turnout, I repeat the estimations of the impact on vote shares but with a different definition of vote shares. I define the latter as the ratio of votes cast to *total registered voters* per municipality. I present the results in table III.34 in the Appendix. I find that the results remain qualitatively and quantitatively consistent with the results observed when vote share (defined as votes relative to presidential votes cast) is the dependent variable, albeit the effect on the former is smaller in magnitude than that of the latter. So the reduction in turnout ratio (lower mobilisation) cannot fully explain the significant increase in the vote share for Alckmin, and that there is evidence of voters *switching* their preferences.[49]

Second, in the following exercise, and in the absence of data on the electorate's preferred candidates, I attempt to decompose the impact observed on vote shares to shed light on whether it is driven by mobilisation, which is the change in the turnout for the election, or by switching votes, voters actually changing their voting towards a different candidate (Neto, 2018; Law, 2021). I perform this analysis for three prominent candidates in this analysis: Geraldo Alckmin, Fernando Haddad, and Jair Bolsonaro.

$$\frac{Votes}{\text{Actual Voters}} = \frac{Votes}{\text{Registered Voters}} \div \frac{\text{Actual Voters}}{\text{Registered Voters}} \tag{6.8}$$

$$\frac{Votes}{\text{Registered Voters}} = \underbrace{\frac{Votes}{\text{Actual Voters}}}_{\text{Vote Share}} \times \underbrace{\frac{\text{Actual Voters}}{\text{Registered Voters}}}_{\text{Turnout Ratio}} \tag{6.9}$$

Therefore, a total differentiation with respect to the ratio of Haitian immigration $dHTI$ gives the following equation:

[49]The impact on vote shares of Bolsonaro and Haddad, defined as votes relative to registered voters, is smaller than that reported for vote shares in table III.9. I also re-estimated equation (5.1) 1- controlling for turnout ratio, and 2- for groups of municipalities with a similar value on the turnout ratio. I find a significant and positive impact on vote shares for Alckmin, with an effect size statistically insignificant from that reported in table III.9. I interpret this result to confirm that the mobilising effect (lower turnout) played less of a role in explaining the increase in the vote shares for this candidate, and that rather results are indicative that voters switched their choices towards Alckmin.

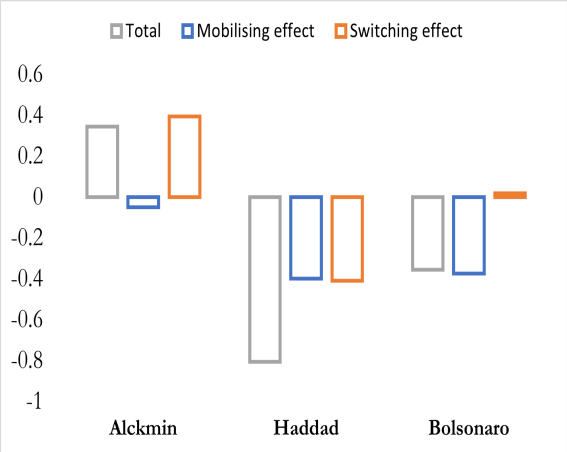
$$\frac{d(\frac{Votes}{\text{Registered Voters}})}{dHTI} = \underbrace{\frac{d\text{Vote Share}}{dHTI} \times \text{Turnout Ratio}}_{\text{Switching Effect}} + \underbrace{\frac{d\text{Turnout Ratio}}{dHTI} \times \text{Vote Share}}_{\text{Mobilising Effect}} \tag{6.10}$$

I justify the interpretation of this decomposition as follows. The first term can interpreted as the switching effect because it measures the change in vote shares for a certain candidate at a given turnout ratio. The second term represents a mobilisation effect because it measures the change in turnout ratio, at a given vote share for a certain candidate. I take the average of turnout ratio and vote shares for each of the candidates across all municipalities in the first round of the 2018 elections. Further, partial derivatives can be directly observed from the estimation of the impact of the Haitian immigration on the respective variables using the 2 SLS estimation of equation (5.1). Then, I calculate the contribution of each mechanism and report the results in figure III.7. I find that, consistent with the indicative evidence reported earlier, that the increase in vote shares for Alckmin is driven by a positive switching effect leading to a 0.39 percentage point increase in vote shares, and a negative effect due to lower mobilisation, by 0.049 percentage points. However, the reduction in vote shares for both Haddad and Bolsonaro was probably driven by lower participation of supportive voters, and, in the case of Haddad, by less voters switching their votes.

This analysis is indicative that there were voters who switched their vote away from Haddad, and towards Alckmin. While Bolsonaro's vote share is mainly impacted by the lower mobilisation (reduction of -0.39 percentage points), despite a very meager switching effect in his favor (0.019 percentage points). With this in mind, I turn my attention next to analysing heterogeneous impacts across municipalities and potential channels that could explain the changes in turnout and vote shares.[50]

[50]In a separate robustness check, I follow more closely the logic behind the decomposition exercise of Law (2021). The only difference in the methodology is that I am able to detect switching effects as the vote share *away* from other candidates. I find the results to be consistent with what I presented in this section.

Figure III.7: Effects of Haitian Immigration on Vote Shares by Mobilising or Switching Effects



Notes: This figure shows the visual depiction of the impact of Haitian immigration on vote shares, defined with as votes relative to registered voters, for each of Alckmin, Bolsonaro, and Haddad, decomposed into mobilising and shifting effects as per equation (6.10). The total impact refers to the impact of Haitian immigration on voting shares defined as the ratio of votes to candidate j as a ratio of registered voters.

6.3 Heterogeneous Impacts

6.3.1 Differential Impacts Based on Baseline Municipality Characteristics

My analysis thus far represents the average effect of Haitian immigration on political outcomes across Brazilian municipalities, which is the most disaggregated administrative level. Those municipalities exhibit heterogeneity in many characteristics that could ameliorate the change in voter behaviour due to the presence of Haitian immigrants. Therefore, in this section, I present evidence on the heterogeneous impacts of the Haitian immigration on turnout ratio and Presidential vote shares for different sub-samples of municipalities depending on baseline characteristics that are theoretically relevant to political behaviour. I also present a brief analysis to see whether the political impact depends on Haitian immigrant visa status. In order to do so, I divide the sample based on these characteristics and re-estimate equation (5.1) on each

sample.[51]

I present the results in tables III.41 and III.42 in the Appendix. In what follows, I discuss the rationale behind the choice of each of the municipality characteristic used and the results associated with each.

Urban and rural: Urban and rural regions may exhibit distinct characteristics that can differentially mediate the impact of the Haitian immigration on voting behaviour. The urban and rural divide tends to also mirror a divide in the economic well-being, available amenities and the subjective evaluation of one's own economic status. A voter, when making a voting decision, can trace the change in their well-being over time, but can also compare their well-being to a certain reference group, for example, their neighbours or people within the same vicinity (Marcinkiewicz, 2018). This ability to compare and observe others, alongside potential cultural divide between regions, makes it useful to study the impact of immigration on voting, subdivided by urban and rural regions. When disaggregating the sample into municipalities belonging to rural or urban regions, I find a decrease in the overall turnout ratio and the turnout ratio of females and males, in response to an increase in the ratio of Haitian immigration to the population, but is insignificant when looking at the urban and rural regions separately. I do not find that voters significantly impact presidential vote share as a result of a change in Haitian immigration to the population.

Large and small areas: The geography of an area, and here measured in terms of size (in km^2), could also be a factor influencing voting behaviour. Saarimaa, and Tukiainen (2016) show how voting can differ depending on the district size because voters in larger areas may feel that their votes influence the elections less, and hence might be less inclined to show up to the voting booths. I define large municipal areas as being all municipalities with an area size between the median and the 99^{th} percentile of my sample, whereas small sized municipalities are those with an area that is at most equal to the median area size. The picture on the impact on turnout ratio is again

[51]Because the estimation is done on different sub-samples, I cannot estimate the impact with precision for some samples as the first stage Kleibergen-Paap F statistic falls below 10 in some cases, so this should be kept in mind when the results are interpreted. I report the Anderson-Rubin test in all the specifications for weak instrumental variable inference. One can also be interested in estimating heterogeneous impacts for the entire sample by employing in the estimation an interaction term between the specific characteristic and Haitian ratio to the population per municipality. However, I am interested in this exercise to study the impact of Haitian immigrants on municipalities that share a certain characteristic. Further, a downside to the interaction method approach is that this will require finding and using a second instrumental variable for the interacted characteristics as many of them are endogenous to the relationship.

different between these two size groups of municipalities. In response to an increase in the ratio of Haitian immigrants to total population, voters in municipalities with larger areas had less political participation (i.e., turnout ratios were lower) in total as well as for males, while voters in smaller municipalities saw a reduction in the turnout ratio in total, and for both genders. The magnitudes of this impact are almost three times larger in large municipalities compared to smaller ones, but the point estimates for large municipalities are statistically insignificant. It is worth noting that measuring area size in terms of population size in 2010 leads to the same results, qualitatively, as reported in this analysis. I find that Haitian immigration impacted adversely vote shares to Bolsonaro in large municipalities, and favourably so in small municipalities. Albeit those results are statistically insignificant. The highly significant increase in vote shares for Geraldo Alckmin is found in the comparison of Haitian-rich municipalities versus less Haitian-rich municipalities within small-sized municipalities.

Inequality: Inequality not only affects voters' perceptions of their relative economic ranking within a society, but also changes people's preferences as they pertain to redistribution policies. The basis of this argument goes back to Schattschneider (1960) who posited the idea that economic inequality is a "form of voter suppression". If the political system is perceived by the poor to work mainly for the advantage of the rich and affluent, then the poor and working class are less likely to vote. Solt (2010) studies how inequality impacted voter participation in the United States across 45 years and finds that there is less likelihood for people residing in states with high inequality to vote. Inequality could also be the outcome of underlying racial inequality in a given municipality. In my analysis, I define low inequality areas as those with a Gini index in 2010 that is less than or equal to the median of the indices of all municipalities; I then study how the presence of Haitian immigrants affects voter behaviour within those areas. I find that there is a negative and statistically significant impact on voter turnout, and there is an increase in vote shares to Geraldo Alckmin in response to an increase in the Haitian immigration ratio.

Pre-immigration Left-leaning municipalities: Voting behaviour can be persistent over time. However, prior to the 2018 elections, there was a huge corruption scandal that implicated, among others, main politicians from the left-wing party (PT). Therefore, municipalities that had majority votes to the left party in the elections prior to the onset of immigration in 2010, might also exhibit a differential response to the

presence of immigrants than municipalities that didn't. I consider those municipalities with a vote share for the left party in 2010 that was at least 50%. I find a reduction in the turnout ratio overall and for both females and males that is significant at the 10% level. I find that voters within this category of municipalities decreased their vote share towards the left-wing, Presidential candidate Fernando Haddad, and increased their vote shares towards Jair Bolsonaro, when faced with an increase in the ratio of Haitian immigrants to the population.

Poverty: Research on voter turnout points to evidence that voters from a poorer socioeconomic background are less likely to vote (Jost, 2021). This observed difference between less affluent and more affluent voters in their political participation can be attributed to individual characteristics such as time availability, lack of resources (as political participation is costly), and civic skills (Armingeon and Schädel, 2015; Johnston and Pattie, 2005; Dalton and Klingemann, 2011; Kern et al., 2015). I study, within the sub-sample of poorer municipalities, how voters altered their political behaviour in response to an increase in Haitian immigrants. This sub-sample of municipalities are those with a ratio of Bolsa Família recipients to population that exceeds the median. Interestingly, within those municipalities, and in response to an increase in the Haitian ratio to the population, I find a significant and negative impact on turnout ratio for both females and males, with large magnitudes relative to the other observed estimates across the different characteristics studied.[52] I find no impact on the vote shares of any of the Presidential candidates.

Migration Concentration: Recent studies have shown that, contrary to conventional wisdom, immigration, may not always trigger right-wing backlash (Steinmayr, 2020; Dustmann et al. 2019). This observation can be explained through the "contact hypothesis" lens (Alesina and Tabellini, 2021), whereby after a period of co-existence, cohesion between groups can start to form and natives can have more positive attitudes towards immigrants. I test this hypothesis by estimating the impact of Haitian immigration on political outcomes for the sub-sample of municipalities that received particularly high numbers of Haitian immigrants, such that they ended up in the top quartile of municipalities for the ratio of Haitian immigrants to population. In those areas, I find that, in response to an increase in Haitian immigration, voters voted less

[52]My results are consistent with using the poverty rate instead of the ratio of Bolsa Família to population.

for Jair Bolsonaro and Ciro Gomes.[53] Interestingly, I find that there is no impact of Haitian immigration on turnout ratio in this sub-sample. However, when I study the impact of immigration within a subsample that has this higher concentration of Haitians and municipalities that have no Haitian immigrants, I find a significant and negative impact on the turnout ratio due to an increase in Haitian immigration. This result could also be lending evidence to the channel of increased heterogeneity due to the presence of Haitians which could lead to dampening political participation (Mutz, 2006).

Differential Impacts Based on Visa Status: Immigrants holding different visa statuses can be different across many observable and unobservable characteristics. A specific visa status determines the extent to which an immigrant assimilates in a community due to rules pertaining to workers' rights, property ownership rights, and rights to social services. In my context, I focus on the two laws that allowed Haitian immigrants and some Asylum seekers to receive the temporary and permanent visa statuses. These statuses reflect mainly the timing of the immigration and the circumstances surrounding the particular visa law. Permanent visas were given with the onset of the crisis in two stages: first, those visas were given through the Brazilian embassy in Port au Prince, with a limit of 100 per month and then, later, through Brazilian consulates in different neighbouring countries, without a limit on monthly visas given. The permanent visa legislation came as a response to the illegal entry of immigrants into the country in large numbers. Temporary visas were granted to Haitian nationals and stateless people residing in Haiti and came into effect on 9^{th} of April, 2018. The latter type of visas can only be issued in Port au Prince. This temporary visa legislation triggered a protest in São Paulo where approximately 200 people, including the current President Jair Bolsonaro, voiced opposition against the immigration law and accused immigrants of bringing terrorists into the country. From a legal perspective, these visas granted identical rights; Both types of visas allow the immigrant to work (Uebel, 2020).

So this analysis aims to understand whether the political reactions of natives depended on the longevity of the exposure to immigrants (another approach to test the contact hypothesis) and the connotations behind the laws involved. By the 2018 elections, natives have been more exposed to the asylum seekers and permanent visa

[53] I also compare municipalities that received more than the median of the ratio of migrants to population with those that received no Haitian immigrants and find qualitatively similar results, albeit insignificant.

holders more than the temporary visa holders. Although Haitians with different types of visas overlapped in terms of their locations across municipalities, however there were areas were the two visa holders located differently.[54] I present the results in tables III.43 and III.44 in the Appendix. I find that the impact on an increase in the vote share is driven by the voters in municipalities that are exposed more to the recent waves of the temporary visa holders. I also find that the negative impact on voter turnout is the same qualitatively between temporary or permanent visa holders, but differs quantitatively. In particular, the impact of temporary visa holders is approximately six times larger than that for permanent visa holders (the estimates are also significantly different than one another at the 10% level). I find no impact on vote shares of turnout ratios due to the presence of Asylum seekers. This could be due to the small numbers of Asylum seekers with an active status in the country.

Haitian Entry Municipalities and Comparison with Venezuelan Immigration: I study whether the main entry municipalities for the Haitian immigrants at the onset of the crisis in 2010 showed a different impact on political outcomes. I perform a comparison of the political outcomes as a result of the Haitian immigrants. I employ a synthetic control analysis to estimate the impact of Haitian immigration on voting outcomes for five municipalities that constitute main entry points for the Haitians (Brasiléia and Rio Branco in the state Acre and Tabatinga in the state Amazonas). In addition, I compare the results with those of the Venezuelan immigration, the most recent and large immigration flow, on Venezuelans' main entry municipalities. As previously discussed, the data on Venezuelan immigrants is limited and mainly shows a concentration of immigrants in the main points of entry (Boa Vista and Pacaraima in the state Roraima). Table III.57 in Appendix section III.F presents the results and details and limitations of this analysis. I do not find that there is any significant impact on turnout ratio, or vote shares to the PT or PSDB parties in municipalities that were initially entry points for Haitians, although the sign of impact on vote shares towards Alckmin is positive. However, I find that there is an increase in turnout ratio and vote share of the PT party in Pacaraima, a main municipality of entry for the most recent Venezuelan immigration. Although initially Haitians arrived to these entry municipalities, after the ease of visa restrictions in 2013 and 2015, they left the original camps and settled in different places across Brazil. In addition, in later years, Haitians started

[54] I find that temporary and resident visa holders located in 47 municipalities that had no permanent visa holders who are Haitian. Whereas the latter located in 56 different municipalities with no permanent or resident Haitian visa holders.

arriving in large numbers into Brazil and through different entry points, including in some cases, air travel. This could potentially explain an insignificant impact on voting behaviour in these municipalities.

6.4 Channels

I have presented thus far evidence on the impact of the Haitian immigration wave into Brazil on electoral outcomes in 2018. I have found that the Haitian immigration had an adverse impact on compulsory turnout across the majority of education levels for both females and males, and resulted in an increase in the vote shares towards more right-wing parties. The impact on the turnout ratio is also found within poorer and smaller municipalities, while the impact on vote shares is found within smaller municipalities and with lower levels of inequality. In this section, I try to understand better what are the channels, at a municipal level, through which Haitian immigration can have this observed effect on political participation and voting behaviour. I divide this analysis into five main categories of the channels of impact: welfare-state, natives' flight, changes in candidate composition, economic, and cultural factors. In what follows, I provide an explanation on the components and the rationale behind these plausible channels of impact.[55] I estimate the impact on the outcomes based on the 2SLS estimation from equation (5.1).

Welfare-state and Native Flight Channels: The hypothesis behind what I term as social channels is that immigrants impact natives' political behaviour through perceived competition over common resources. This can be particularly true for immigrants arriving due to a exogenous push factor and thus might require an increase in fiscal resources to accommodate for their presence (which may not be the case for economic immigrants). Haitians fit this category because they initially arrived due to a devastating earthquake in Port au Prince and needed fiscal support in the host communities they went to. I explore whether the observed outcomes can be rationalised through the natives' internal migration due to the presence of Haitians. This factor will impact the turnout ratio, as well as voting shares if those who migrated are of a particular political affiliation. I present evidence to understand whether my observed

[55] Although the following results are estimated for the universe of municipalities, I conduct as a robustness check a similar analysis of the channels but based on sub-samples of municipalities identified in section 6.3. I find the results to be consistent with the average impacts reported in this section, 6.4.

political outcomes are due to natives' flight (or voting with their feet). In order to do so, I create an internal migration measure for each municipality that represents *net* yearly internal migration of natives as a proportion of population in 2018. I describe the calculation of this measure in equation (6.11).[56] . Natives might also respond to an increase in local fiscal support by predicting an increase in future taxation, or a decrease in welfare benefits transferred to them and their families. I study the impact on six social outcomes as potential channels through which immigrants influence natives' political behaviour. The variables I use proxy for three overarching topics: *Fiscal commitment*, measured by the collected tax values per municipality, social expenditure, agriculture expenditure, and the overall budget expenditure in 2018 weighted by GDP in 2010; *Poverty* proxied by the number of Bolsa Família beneficiaries and the amount of Bolsa Família transfers in 2018 (weighted by population size and GDP in 2010, respectively); and *social unrest* proxied by homicides in 2018 weighted by population size in 2010.[57]

I present the results of the impact on social outcomes in table III.10. I find no significant impact of Haitian immigration on any of these outcomes except for a statistically significant decrease (at the 10% level) in local agriculture expenditure. A one percentage point increase in Haitian immigration relative to the population decreased the ratio of agriculture expenditure in 2018 relative to GDP in 2010 by 0.2 percentage points.

$$\frac{InternalMigration}{Population_{t-1}}_{m,t} = (Pop_{m,t} - Pop_{m,t-1}) - \Big[(births_{m,t} - deaths_{m,t}) - (births_{m,t-1} - deaths_{m,t-1})\Big] \tag{6.11}$$

Candidate Composition Channel: It could also be the case that the presence of Haitian immigrants influences the type of candidates running for election which might in turn influence voters' behaviour.[58] If the presence of immigrants in a particular municipality alters the willingness of certain individuals to run for elections, then this

[56] I also explore three different measures on the natives' internal migration and I describe and present the results in appendix III.E.3.1. I do not find any impact on natives' flight due to the presence of Haitian immigrants.

[57] Poverty rate at the municipality level is not available for 2018 from the *O Instituto Brasileiro de Geografia e Estatistica, IBGE.*

[58] This analysis would have been more apt if studied at the level of municipal elections, which goes beyond the scope of this paper. I still believe that even if I am studying national level elections, immigrants can potentially impact candidates willingness to run for elections, depending if their municipality of birth received large influxes of immigrants relative to other municipalities.

Table III.10: Channels of Impact: Welfare-state and Native Flight Outcomes

Social Outcomes								
	Natives' Flight	**Fiscal Commitment**				**Poverty**		**Social Unrest**
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	$\frac{InternalMigration}{Pop_{t-1}}$	Tax Values	Social Expenditure	Agriculture Expenditure	Budget Expenditure	Number of Bolsa Familia	Value of Bolsa Familia	Homicides
HTI Flows (interval)/Pop 2010	0.00116	0.206	-0.000146	-0.00223***	0.00406	-0.00615	0.000408	0.0000251
	(0.00283)	(0.000131)	(0.00143)	(0.000593)	(0.00895)	(0.0266)	(0.00132)	(0.0000269)
Mean of Outcome Variable	-0.00473	0.145	0.0291	0.00833	0.285	1.210	0.0430	0.000223
Mesoregion F.E	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The table shows the second stage 2SLS estimates for the impact of migration on different social outcomes. Each outcome variable represents the value of this outcome in 2018, weighted by population in 2010 for homicides and number of Bolsa Familia, and weighted by GDP in 2010 for the rest of the variables. The sample size varies depending on the availability of the outcome variable in the baseline year of 2010. Equation 5.1 is estimated for each of these outcome variables. Regressions include mesoregion fixed effects and municipality controls in 2010. Results are shown for the full sample of municipalities for which data on control variables is available. The predicted Haitian migration variable is adjusted by population in 2010. The mean of the dependent variable is shown for municipalities that had a Haitian immigrant in 2018. Data sources: IBGE 2010 Census data for control variables, Tribunal Superior Eleitoral and STI-MAR and SISMIGRA for migration data.

could impact voter composition, and in turn affect voter behaviour. This hypothesised channel is closely related to the work done by Adams and Merrill III (2003). The authors argue that non-policy variables, such as a candidate's socio-demographic characteristics can affect voters' political participation and voting choices. Therefore, I link each candidate running for Federal Deputy, State Deputy, Governor, and President, in the 2018 elections to their municipality of birth and estimate whether the presence of Haitian immigrants in those municipalities impacted the candidate characteristics and numbers running for elections. I present the results in table III.53 in the appendix on total number of candidate and across different candidate characteristics such as their race, education and gender. I find no statistical impact on any of candidate composition characteristics due to having received Haitian immigrants in their municipality of birth. Therefore I can rule out this channel.

Economic Channel: Having presented the results on the social and welfare channel, I study next if immigrants impacted voting behaviour via the economic channel. This channel, with a focus on the impacts in the labour market, is one that has been studied at length in the immigration literature, and it states that immigrants can have positive impacts on wages and employment opportunities of complementary types of workers but may affect adversely the labour market outcomes of substitute workers. Therefore, the underlying hypothesis is that natives will change their voting behaviour

based on whether or not immigrants affect their relative socio-economic ranking in the society or their labour market prospects. I present evidence on this channel of impact across eight different variables measured at the municipal level: *average economic outcomes* is measured by employed people, paid employees, average salary and number of companies; *Value Added by Sector* measured by value added in agriculture, industry and services. I present the results in table III.11.[59] I find that the presence of Haitian immigrants increased the number of companies per population in 2018, and the value added in services and industry to GDP in 2018 due to an increase in the ratio of Haitian immigrants to the population. For example, a one percentage point increase in the Haitian immigration ratio leads to a 0.04 percentage point increase in ratio of the number of companies to the population. I note that such impacts are also economically large relative to the mean of the channels studied, which is surprising given the small number of Haitians relative to the Brazilian population.[60] This is because the resulting impacts are due to a one percentage point increase in the Haitian ratio to the population, which represents approximately a 42% increase from the mean (0.022%) of this variable. Another way of understanding the results is that an increase by the mean of the ratio of the Haitian immigration to the population, which would be equivalent to a doubling of this ratio across municipalities, would lead to a 0.0062 and a 0.0032 percentage point increases in the ratio of the value added of industry and services to GDP, respectively. Although statistically significant, perhaps reflecting their contribution to the formal labour market, these effects are economically modest, potentially due to the small number of Haitian immigrants relative to the Brazilian population.

Cultural Channel: Next, I turn my attention to the role of group-related concerns as another potential channel of impact.[61] Those concerns include xenophobia, identity

[59]Ideally, a detailed analysis at the worker level on the formal and informal labour market channel can be done where the researcher can study the impact for workers at different skill levels in order to make a statement about the complementarity and substitutability between natives and migrants. This goes beyond the scope of this paper. At this stage, I am unable to draw such conclusions as I explore this impact at a municipal level. I acknowledge that the impacts at the municipal level may have different distributional consequences for natives. Currently, I lack the data, especially for informal workers and 2018 formal workers, to explore this channel at such a detailed level, which remains a limitation in this analysis, however this is a topic that I plan to undertake in future work.

[60]Those effects could also driven by small municipalities; therefore investigating how the impacts on those channels differ based on municipality characteristics is a future avenue for research.

[61]Note that this section presents descriptive analysis to shed light on some of the channels related to individual preferences. The reported impacts represent average effects and they are not directly comparable with the local average treatment effects of the 2SLS estimations presented earlier.

Table III.11: Channels of Impact: Economic Outcomes

Economic Outcomes							
	Average Economic Outcomes				**Value Added by Sector**		
	(1) Employed People	(2) Paid Employees	(3) Average Salary (Reais)	(4) Number of Companies	(5) VA Agriculture	(6) VA Industry	(7) VA Services
HTI Flows (interval)/Pop 2010	0.0341	0.0268	70.53	0.00432*	-0.0144	0.282**	0.148**
	(0.0254)	(0.0247)	(64.06)	(0.00246)	(0.0149)	(0.143)	(0.0668)
Observations	5558	5558	5558	5558	5558	5558	5558
Mean of Outcome Variable	0.152	0.127	1926.8	0.0198	0.379	0.294	0.655
Mesoregion F.E	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The table shows the second stage 2SLS estimates for the impact of migration on different economic outcomes. Each outcome variable represents the value of this outcome in 2018, weighted by population in 2010 for Employed People, Paid Employees, Salary, and number of Companies, and weighted by GDP in 2010 for the rest of the variables (Value added in agriculture, industry, and services). The sample size varies depending on the availability of the outcome variable in the baseline year of 2010. Equation 5.1 is estimated for each of these outcome variables. Regressions include mesoregion fixed effects and municipality controls in 2010. Results are shown for the full sample of municipalities for which data on control variables is available. The predicted Haitian migration variable is adjusted by population in 2010. The mean of the dependent variable is shown for municipalities that had a Haitian immigrant in 2018. Data sources: IBGE 2010 Census data for control variables, Tribunal Superior Eleitoral and STI-MAR and SISMIGRA for migration data.

threats, or general worries about having a different cultural group within the same vicinity as the voter themselves.

In order to present insights on this channel, I analyse the Latin American Public Opinion Project dataset for 2019 collected by Vanderbilt University. The analysis presented herein is descriptive in nature and at the individual level. This dataset has interviews for over a 1,480 individuals and is nationally representative. What is interesting about this specific wave is that it contains questions regarding attitudes towards immigrants and questions on the individual's political behaviour and preferences in the Presidential election of 2018.

First, in table III.12, I present descriptive statistics on the percent of respondents who voted for a certain presidential candidate, and by their political participation choice (election turnout) in table III.52, disaggregated by their preferences on immigration. I do not find that there is a large difference in the percentage of people who report "strongly agreeing" to offering social services for immigrants versus those that report "Somewhat disagreeing" in terms of political participation. I find that 37.5% of those who answered by "a lot" to being bothered by having an immigrant as a neighbour, report not participating in the elections in 2018. Non-participation is highest among other groups of voters who do not report being as bothered by an immigrant as a neighbour (those results are reported in table III.52 in the appendix). I report in

the table that 63% of those "strongly disagree" to offering immigrants social services, also report to have voted for Bolsonaro in 2018. Interestingly, I find that those who do not object to having immigrants as a neighbour (chose "Nothing" or "Little" to being bothered by an immigrant as a neighbour) voted for Bolsonaro as opposed to Haddad. This could be reflecting two things: the first is simply that the sample has more voters that voted to Bolsonaro, which is a likely case considering the latter won the election. The second, if we believe that the survey is nationally representative, is that Bolsonaro voters do not seem to mind having immigrants as neighbours, but object to offering social services to those immigrants.

Table III.12: Descriptive Statistics on Presidential Voting Choices by Individual's Preferences on Immigration Policies

	Who did you vote for in the last presidential election of 2018?											
	Blank Vote	**Didn't Vote**	**Jair Bolsonaro (PSL)**	**Fernando Haddad (PT)**	**Ciro Gomes (PDT)**	**Geraldo Alckmin (PSDB)**	**João Amoêdo (NOVO)**	**Cabo Daciolo (PATRI)**	**Marina Silva (REDE)**	**Álvaro Dias (WECAN)**	**Other**	**Total**
	%	%	%	%	%	%	%	%	%	%	%	%
Should government offer social services to immigrants?												
Strongly agree	26.3	36.8	34.7	41.2	38.6	73.9	40.3	0.0	0.0	100.0	45.7	36.6
Somewhat agree	33.1	32.6	35.6	27.6	36.2	0.0	23.6	0.0	0.0	0.0	54.3	32.9
Neither agree nor disagree	25.7	26.4	8.2	8.9	8.7	26.1	20.7	100.0	100.0	0.0	0.0	11.0
Somewhat disagree	0.0	4.2	9.8	12.4	10.7	0.0	0.0	0.0	0.0	0.0	0.0	9.4
Strongly disagree	14.8	0.0	11.6	10.0	5.7	0.0	15.3	0.0	0.0	0.0	0.0	10.2
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
How much would it bother you to have an immigrant as a neighbor?												
a lot	0.0	0.0	1.9	6.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.5
Somehow	9.7	4.2	1.3	2.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.9
Little	12.0	20.8	10.4	8.3	5.7	0.0	20.7	0.0	0.0	0.0	28.9	10.3
Nothing	78.3	75.0	86.3	83.1	94.3	100.0	79.3	100.0	100.0	100.0	71.1	85.3
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Notes: The table shows descriptive statistics based on the Presidential voting choices of Brazilians in 2018 as well as their attitudes towards immigrants. The survey asks directly about Venezuelans, Spanish and all other immigrants (not specifically about Haitians). The numbers in the table represent the percent of people, weighted by the sampling weights, that belonged to each category. Source: LAPOP dataset 2019.

First, I assess which individual preferences are correlated with a person's decision to vote or abstain. I am interested in how those preferences differ between in Haitian-rich municipalities and those municipalities that didn't receive Haitians. I estimate the following equation:

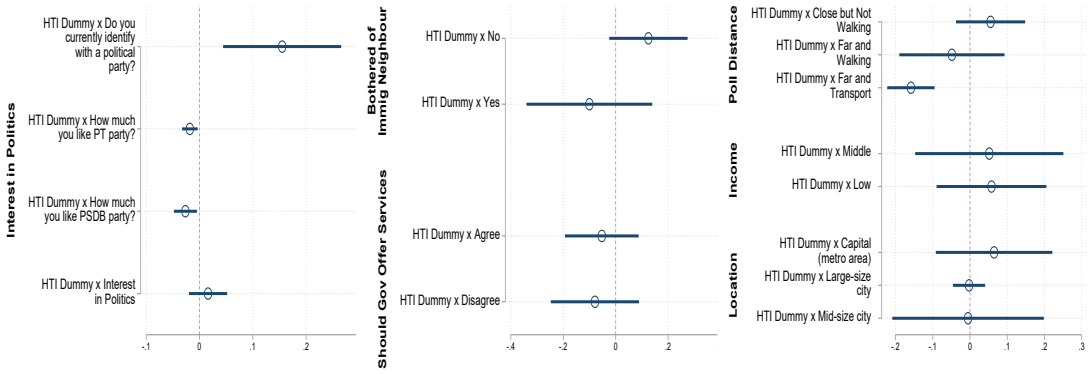
$$Turnout_{ind,i} = \alpha_{ind,i} + I(MIG_i = 1) \times Y_{ind,i} + \chi_{ind,i} + \epsilon_{ind,i} \tag{6.12}$$

where $Turnout_{ind,i}$ is a dummy variable indicating 1 if the individual voted in the 2018 elections. I find evidence of three results. I plot the point estimates and the confidence interval at the 10% level in Figure III.8. The first significant result shows that those individuals in Haitian-rich municipalities who think that the government should not offer social services to immigrants, were less likely to participate in the 2018 elections. Further, I find that voters in municipalities that received Haitians were

less likely to participate in the elections if they lived far from the polling station and required transport.

Figure III.8: Results on Individual Turnout

(a) Based on Political Preferences (b) Based on Immigration Preferences (c) Based on Individual Characteristics



Notes: The figures show the results from estimating equation (6.12). The three graphs show results from separate estimated equations to maximise the number of respondents in the each of the estimation. All estimations control for individual level characteristics such as age, sex and area. The point estimates represent the estimates from the interaction of each variable by a dummy indicating whether the municipality has Haitian immigrants or not. Confidence intervals at the 90% level are depicted around the point estimates.

Another factor that could impact voter turnout is whether or not the voter believes the candidates have the power, ability and interest to address the issues they care most about. Therefore, I estimate the correlation between what voters believe to be the most important issues with whether or not a Haitian immigrant is present in their municipality and I present the results in figure III.9(a). Then, I estimate equation 6.12, where $Y_{ind,i}$ represents a person's choice of what they consider to be the most important problem facing Brazil currently. I present the results in figure III.9(b). I find that municipalities that have at least one Haitian immigrant, also have voters who believe that discrimination, street gangs, armed conflict, crime and transport were the main issues facing the country.[62] However, the perception of the individual's problem translates into different political choices. In particular, I find that individuals who believe that discrimination is the most important issue are less likely to participate in the elections if they lived in an area where Haitian immigrants located. This result is opposite for those who believe that drugs and that the government is "evil". The latter

[62]Discrimination in the survey is presented as a general definition. It doesn't refer specifically to whether discrimination is based on race or nationality.

set of voters were more likely to show up to the voting booths if they were in Haitian abundant areas versus areas that didn't receive any Haitians.

Figure III.9: Correlation of Opinions on the Most Important Problem with Haitian Presence and Voter Turnout

(a) Correlation with Haitian Presence

(b) Correlation with Voter Turnout

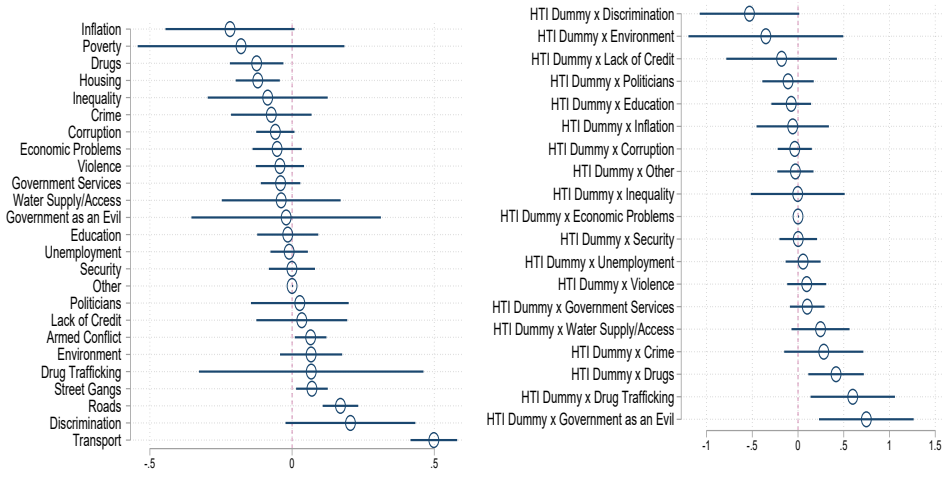


Figure III.9(a) shows the correlation between what individuals think is the most important problem facing their country with the presence of Haitians in their municipality. Figure III.9(b) shows the results from estimating equation (6.12). The point estimates represent the estimates from the interaction of each variable by a dummy indicating whether the municipality has Haitian immigrants or not. Confidence intervals at the 90% level are depicted around the point estimates.

In order to assess social attitudes in relation to the presence of Haitian immigrants in a certain municipality, I link the migration data to the LAPOP sample of municipalities. I estimate the following multinomial logit estimation:[63]

[63]This analysis remains descriptive in nature and has caveats. While it would be ideal to identify the impact on attitudes using the IV strategy employed earlier, the LAPOP dataset's small sample of municipalities and individual level sampling does not allow for that. Furthermore, the assumptions surrounding the multinomial logistic regression can be restrictive because it assumes that the error terms are independently and identically distributed which may not be the case here. Another main assumption is the independence of irrelevant alternatives (IIA). In a political voting context, this assumption can be violated if the introduction of a new candidate (C) changes the relative choice probability between candidates A and B. I believe that in my context, the removal or addition of alternatives will vary the odds of the other alternatives in the same proportion and each of the candidates stand for different ideological positions. I test this assumption using the Hausman-McFadden test (1984) and do not find evidence that the IIA has been violated. However, Long and Freese (2014) show that many of the tests used for IIA assumption tend to produce conflicting results and do not encourage their use. To address this issue, future work will investigate how the estimates from the multinomial logit compare to the nested logit and multinomial probit choice model. That said, I attempt a stratified approach based on ideological characteristics and find the results to hold to what is reported in this section.

$$Y_{ind,i} = \alpha_{ind,i} + \frac{HTI}{Pop}_{i,2018} + \chi_{ind,i} + \epsilon_{ind,i} \tag{6.13}$$

where $Y_{ind,i}$ is an ordinal variable that represents a person's *ind* preference over a certain question, including their presidential candidate choice, in municipality i; χ_{ind} is a vector of individual level characteristics[64] and $\epsilon_{ind,i}$ is an error term. I plot the probability results in figure III.10 across different values of the ratio of the Haitian immigration to population (%).

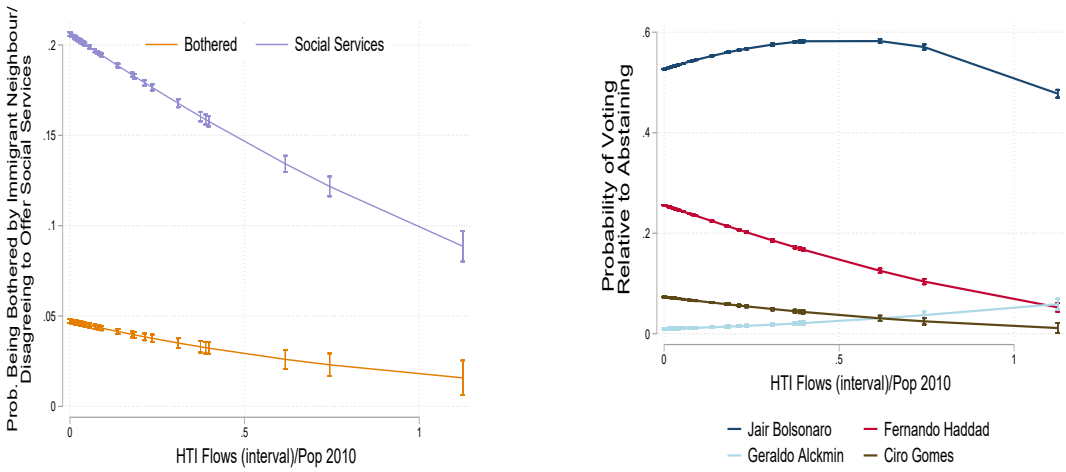
I find that the probability of being bothered by an immigrant as a neighbour or the probability of disagreeing to offer social services to an immigrant (relative to being indifferent) decreases with the increase in the concentration of Haitian immigration to population. This finding could be supportive of the *contact hypothesis*. Further, I estimate the probability of voting for either Fernando Haddad (the PT party), Jair Bolsonaro (the PSL party), Geraldo Alckmin (the PSDB party) or Ciro Gomes (the PDT party), relative to those those who abstained. I find that the probability of voting for Jair Bolsonaro or voting for Geraldo Alckmin, relative to not voting, increases with the concentration of Haitian immigration per population; but for Bolsonaro, the probability decreases after a certain threshold of the Haitian migration concentration (at approximately 0.6%, which is approximately higher than the 75^{th} percentile of the Haitian immigrant ratio to population distribution across municipalities). However, the probability of voting for Fernando Haddad decreases relative to abstaining from voting with the increase in the share of the Haitian immigration to the population.

[64]The individual level characteristics I control for are age, gender, income level and type of residential area (rural, and size of city).

Figure III.10: Results from Multinomial Logistic Regression

(a) Probability of Being Bothered *a lot* by an Immigrant and the Probability of Disagreeing to Offer Social Services

(b) Probability of Voting to a Certain Candidate Relative to Abstaining



Notes: The figures show the results from a multinomial logistic regression as per equation (6.13). The probability in either graph varies by the size of the Haitian ratio to the population. Graph III.21(a) represents the probability of stating that the individual is bothered by having an immigrant as a neighbour (blue) and the probability of disagreeing to offer social services to an immigrant (dashed red). The probability varies by the ratio of Haitian ratio to the population. Graph III.10(b) depicts the probability of voting to the top four Presidential candidates at different values of the ratio of Haitians to the population.

6.5 Discussion

My core results can be summarised as follows: an increase in the ratio of Haitian immigration to the population decreases voter turnout (electoral participation),for both females and males and at different education levels, and increases votes for the right-leaning Presidential candidate Geraldo Alckmin. I further provide evidence that those impacts are heterogeneous across different sub-samples of municipalities based on certain characteristics, especially the level of inequality, municipality size, poverty levels and immigration concentration. In this section, I synthesise the findings and propose explanations on the channels of impact.

6.5.1 Impact on Political Participation

The second theme in my findings relates to the impact that Haitian immigration had on voter turnout. I find this result to hold across different education groups, gender and age categories while being mainly concentrated among those for whom it is

compulsory to vote. This result holds despite controlling for a rich set of underlying electorate characteristics (in 2014), among other baseline municipality characteristics. A disproportionate voter participation is problematic for the democratic representation and leads to unequal representation among the interests of the electorate.

Cultural attitudes In addition, I explore whether cultural attitudes affected voter turnout. I provide descriptive evidence that offers two potential explanations. First, I find that voters who resided in areas that received a Haitian immigrant and also required transport to reach the electoral poll, were less likely to participate in the elections. And second, I find that what people believe to be the most important problem facing their country differs in Haitian rich areas versus others and that in turn affects electoral turnout. As shown in the descriptive analysis in III.9, individuals stating that one of the most important problem as "discrimination" were highly correlated with areas that received immigration. Further, I find that those who believe that discrimination is the most important problem and reside in Haitian-rich municipalities were also less likely to show up for voting (significance at the 10% level). I hypothesise that this finding is related to theories of voter uncertainty towards candidates where voters in Haitian-rich areas are not certain which of the candidates is best positioned to address this particular problem, and so the voter is indifferent among the candidates, and therefore, abstains from voting. The impact of uncertainty about candidates on turnout ratio has also been documented in the literature. Adams, Dow and Merrill III (2006) find that *alienation* from the candidates and *indifference* between the candidates led to significant increases in voter abstention. Sanders (2017) argues that uncertainty around candidates can decrease turnout among voters with weak preferences but increase it among voters with strong preferences.

Increase in Heterogeneity or Perceived Relative Ranking in a Community I find evidence of a decrease in political participation in small municipalities, and those where in the baseline had higher levels poverty or those that had lower inequality. The presence of Haitian immigration could increase the perception of heterogeneity, or relative ranking of an individual in the socio-economic ladder.[65] Political theory on determinants of voter participation states that turnout tends to be lower for less well-off individuals who feel that they have less influence on political outcomes, observe larger

[65]I expect this factor to be more pronounced within smaller municipalities and those where initially inequality was low. The Haitian immigration shock could have changed how people perceive themselves relative to others in the socioeconomic ladder.

inequality within their spatial area, who face relatively larger costs and lower incentives to participating in the elections, and who face larger uncertainty about short term choices that long-term choices such as choosing the next candidate doesn't take priority (Denny, 2017; Haushofer and Fehr, 2014; Armingeon and Schädel, 2015; Gallego, 2007). If Haitian immigration helped to heighten those factors or to make individuals more aware of their socio-economic relative position to others in their community, then this can potentially explain the lower voter turnout. I cannot study the differential impact by income level due to lack of data availability at the electorate level, but I proxy for income by the education level, where I find that those with lower levels of education show a slightly larger decrease in turnout ratio than their more educated counterparts, and in the case of males, the reduction in turnout ratio was driven by those with lower levels of education. Further, I do not find any significant impact on changes in the number of Bolsa Familía beneficiaries nor the value of Bolsa Familía per individual. Therefore, there is no evidence that Haitian immigration is impacting political participation through affecting actual poverty or inequality levels, but rather it could be impacting voters' perception of their relative ranking in the community and the level of heterogeneity therein.

Further, Soss and Jacobs (2009) argue that large differences between spatial units can produce "distinct and homogeneous clusters of unequal social privilege and political influence, intensifying the social and political isolation and polarisation of neighbourhoods". So a potential channel of what could be at play, is that those voters in municipalities that received Haitian immigrants, and those that were already more unequal and had higher poverty rates, found themselves to be at a relatively stark difference from other municipalities and that contributed to a sense of political apathy and disengagement, leading to a reduction in the turnout ratio. Smith and Mckie (2000) argue that being exposed to an "out" group solidifies the identity of the "in" group. In areas that became heterogeneous because of the presence of Haitian immigration, a sense of an in group identity can become heightened across all factions of the society. This might create a sense of homogeneity among the non-Haitian group, and reduces the competition within this group, resulting in a demobilising effect. I would expect this "in" group to be more prominent in smaller sized municipalities, and for the impact on turnout to be more pronounced within such municipalities. This is indeed what is reported in the results. This is also consistent with theories that perceived reduction in competition, reduces voter turnout (Pacheco, 2008). When I study the impact of immigration within a sub-sample that has this higher concentration of Haitians and

municipalities that have no Haitian immigrants, I find a significant and negative impact on the turnout ratio due to an increase in Haitian immigration. This result could also be lending evidence to the channel of increased heterogeneity due to the presence of Haitians which could lead to dampening political participation (Mutz, 2006).

6.5.2 Impact on Vote Shares

I start with the observation that Haitian immigration increased votes towards Geraldo Alckmin, who is considered a right-leaning centrist candidate. The presence of Haitian immigration mostly had no impact on vote shares for Fernando Haddad and Jair Bolsonaro, except for the sub-sample of municipalities that had a majority left party win in 2010. I rationalise this impact along three dimensions which are consistent with the hypotheses validated empirically in other contexts: First, the perceived anti-Haitian immigration attitude that Geraldo Alckmin took at the peak of the Haitian immigration, second, his conservative, right leaning political stances and ideologies and, third, evidence for the "contact hypothesis".

Anti-immigration attitudes First, Geraldo Alckmin's role during the Haitian influx into Brazil was not neutral. In 2013, the state of Acre which was receiving a lot of Haitian immigrants declared a "situation of social emergency" by the then governor Tião Viana, who belonged to the PT party (Marcel and Natani, 2013; Damé, 2013). Triggered by the floods that occurred in Acre in 2014, the state closed the shelters and sent about 800 Haitians by bus to the city of São Paulo. This situation triggered a political backlash and led to a widely publicised dispute in the receiving city, with wide media visibility across the country. Back then, Geraldo Alckmin was governor of São Paulo and he objected to the arrival of Haitians, calling the state of Acre as "irresponsible" and accusing it of sending the Haitians without a warning. Since Geraldo Alckmin has been painted as having anti-immigration attitudes by the vast media attention that this incident received, a potential explanation to the increase in the vote share to this candidate in response to Haitian immigration is native backlash to the presence of immigrants. This is a finding that is consistent with other work on the impact of immigration on political outcomes (Tabellini, 2020; Alesina and Tabellini, 2021).

Right-wing ideological beliefs In terms of ideological beliefs, Geraldo Alckmin, as I have depicted in figure III.5 based on Wiesehomeier, Singer, and Ruth-Lovell (2021),

promoted cutting public services in order to cut taxes and advocated the view that transfers are not for everyone and should be dispersed on a discretionary basis. He also opposed redistribution and favoured market regulation whenever possible. On these issues, Alckmin held more conservative positions than Bolsonaro did, and more than any of the other left-wing candidates. So natives might be voting more towards Alckmin either out of fear that taxes might rise or to create some form of native exclusivity over the social transfers (to deny immigrants social transfers). I do not find empirical evidence that increases in the ratio of the Haitian immigrants to the population led to changes in social transfers. One exception is the significant reduction in municipality level agriculture expenditure in response to a one percentage point increase in Haitian immigration. I interpret this result as being consistent with the hypothesis that a potential channel for the observed impacts is citizens' preferences to choose candidates that are conservative in their view of fiscal redistribution, such as Alckmin.

Furthermore, Geraldo Alckmin was seen as a business friendly figure and one who made economic reforms central in his campaign, pointing to his success in the economic performance of São Paulo, where he was governor. So the positive impact I find due to Haitian immigration on the number of companies, and the value added in services and industry, lends support to the idea that voters in areas that received Haitians could see in Alckmin the potential to keep the economic prospects positive in a way that will absorb the new immigrant labour force.

Contact hypothesis In addition, my results lend support to the "contact hypothesis" (Alesina and Tabellini, 2021). I find that voters present in municipalities with high levels of Haitian immigration had no impact on the vote shares to Geraldo Alckmin, and a statistically significant reduction in the vote shares for Jair Bolsonaro. One potential explanation for this result that goes against the held conventional wisdom is that of the contact hypothesis which states that sustained contact with a different group of people who are not perceived as a threat can begin to create cohesion among the groups in the society (Allport, 1954 as cited by Alesina and Tabellini, 2021). This result has also empirical evidence in other contexts. Steinmayr (2020) finds that municipalities where residents had only a transient and short-lived interaction with refugees in Austria voted more towards right-wing parties as opposed to other municipalities where residents had a prolonged interaction and where the refugees settled and sought asylum. Dustmann et al. (2019) finds that in response to refugee inflows in Denmark,

the vote shares towards the right and centre-right parties increased. However, this impact wasn't true for large and urban municipalities where residents voted more towards pro-immigration candidates.

7 Conclusion

The fear of immigration is poisoning Western politics. Donald Trump owes his job to it. Brexit would not be happening without it. Strident nationalists wield power in Italy, Hungary, Poland and Austria, and have gained influence elsewhere.- The Economist, 2018.

International immigration remains a salient and important political discussion. Its high salience presence in western politics has led many researchers to investigate the extent of this impact on voters' behaviour and electoral outcomes. In this article, I depart from the western focus of the impact of immigration, and study the south-south migration context of a large Haitian immigration flow into Brazil in recent years. I study this impact in the unique election of 2018, where the far-right candidate, Jair Bolsonaro, ultimately rose to power. Bolsonaro was vocal about his opinions towards immigrants, and his campaign rhetoric had an anti-immigration stance. He had called Haitian, Senegalese, Bolivian, and Syrian migrants and refugees "the scum of the world" (Espinoza and Brumat, 2018).

I study the political impact of Haitian migration into Brazil by studying the differential impact on municipalities with different numbers and proportions of Haitian immigrants. To establish a causal link, I exploit the spatial variation in the locations of Haitian immigrants and the exogenous earthquake that triggered an influx of Haitians. I then use a variation of the shift-share past settlement instrumental variable strategy to estimate the causal impact of immigration on political outcomes. I look at impacts in terms of voter participation and vote shares. My study is enriched by a disaggregated analysis on the impact of voter turnout by education, gender and age groups. I then study how the impacts vary within subgroups of municipalities sharing certain key characteristics. To complete the analysis, I assess five potential channels that can be at play and driving my results: welfare-state, economic, natives' flight, candidate composition and cultural attitudes. For the latter, I use the 2019 LAPOP dataset to provide descriptive evidence on channels pertaining to opinions formed towards immigrants.

Throughout this paper, I discuss the assumptions behind the instrumental variable empirical strategy employed and I present a series of robustness checks to address some of those concerns. The results presented in this paper are applicable within the Brazilian context and I argue that my empirical strategy provides a reasonable level of internal validity and hence the results can offer potential policy implications. First, I find that voter absenteeism is an important effect of immigration and can affect individuals differently depending on their level of education, age or gender. This impact has not been explored as much in the literature. Reduction in political participation is a problematic finding for the democratic process and for the equal representation of electorate interests. My analysis doesn't indicate that this is driven by economic disadvantages posed by the immigrants, and is rather driven by cultural attitudes towards the presence of an "out" group. Therefore, state interventions or non-governmental targeted campaigns that help keep the electorate engaged can mitigate the impacts of the reduction in political participation. Second, even if immigration constitutes a small proportion of the population (in this context, less than 1%), it can still alter voters' electoral behaviour. And third, although I cannot draw conclusions on the complementarity and substitutability between immigrants and natives in the labour markets, my analysis at the municipality level indicates that the presence of the Haitian immigrants has, if anything, resulted in positive economic outcomes. Since Haitian immigrants were allowed to work by the Brazilian law, this evidence points to the importance of considering inclusive immigration policies that promote effective labour market assimilation of immigrants.

This setting analysed a case of a new migration wave into Brazil that was triggered by an exogenous push force, and therefore broader lessons from such quasi-experimental analysis must be drawn with care. Although Haitian migrants can be considered as economic migrants, Haitian immigrants may not be representative of the typical economic migration waves that are voluntary and pre-planned (Peri, 2016). The latter includes immigrants who may hold job positions prior to arriving to the country or are in touch with a wide network of individuals in the host country. Further, the more common scenario for migrants escaping natural or human-made disasters is to move to the neighbouring countries that may have been affected in other ways by the shocks driving the migration movements. Thus, the Haitian migration to Brazil may be unusual, relative to many episodes of forced migration. The closest evidence we have on this is that of Jaupart (2018) who studies the impact of the Haitian immigration on political outcomes in the neighbouring Dominican Republic. Although my findings on

an increase in vote shares to right-wing leaning candidates is consistent with Jaupart (2018), I find further evidence that there exists a significant reduction in natives' turnout ratio in Brazil in response to the Haitian immigration[66].

Future work on this topic could extend this analysis. For one, studying the economic channels of immigration on political outcomes through the labour market channel is promising. In my future work, I will study how the economic impact will differ based on the educational composition of the workforce. This is in line with the theory that immigrants can be both substitutes and complements to the native workforce; on net, the impact is not clear a priori (Ottiviano and Peri, 2012). In turn, the skill level of the immigrant will have a different impact on the voting behaviour of voters depending on whether or not the latter face competition in the labour market, in both the formal and informal sectors. My current dataset on migration is not adequate for this purpose because the only measure I have to proxy skill level is a noisy indicator on the immigrants' reported profession. More generally, even if the focus is one election year only, the study of the channels of impact can be studied across time and a dynamic analysis on the impact of immigrants on the channels of impact in the short and long term can provide a more complete picture to explain the mechanisms of voter behaviour. Further, the analysis should go beyond the impacts on voting outcomes, to study policy choices and their impacts on future socio-economic policies. The local management of the pandemic at a municipal level might make an interesting case study.

In terms of the empirical strategy, it is promising to study what impacts the location decisions of immigrants, beyond the commonality of the country of origin, to include language and other cultural characteristics. The analysis in this paper exploits language proximity between recent immigration locations and earlier migrant settlements to use more of the historic migrant spatial distribution. But future work that involves experimentation with different methodologies in the calculation of the language proximity index (for example, accounting for different languages spoken within a country and weighting that by the concentration of emigrants from certain regions within the country), as well as a more fledged analysis to understand the contributions of the underlying language shares (similar to what has been presented in this paper) will be essential to test the robustness and validity of this strategy when compared with the country of origin past settlement instrument. Additionally, my analysis shows that

[66]Jaupart (2018) also finds a reduction in the turnout ratio in response to the Haitian immigration presence, that is similar in magnitudes to my reported estimates, albeit the effect is insignificant from zero.

the impact of Haitian immigration was mainly insignificant on the vote shares of the far-right candidate Jair Bolsonaro and was rather directed towards a candidate who was involved in an anti-Haitian scandal. This effect can be explained by understanding further the main pillars of the populism of Bolsonaro and what makes it different than the right-wing populism observed in the west. In addition, there is room to improve the analysis by including the impact of the Haitian immigration on voting outcomes for the Presidential elections across different years[67]. For example, studying the upcoming 2022 elections, will offer new insights regarding the long-term impacts of immigrants. In the same spirit, studying how voters respond to immigrants' presence across different levels of elections in a federal system can allow the researcher to bring to light deeper insights on voter behaviour in terms of the latter's punishment or reward to relevant politicians based on the latter's respective responsibilities. And finally, the analysis in this paper remains empirical in nature, and therefore the answers it provides can be limited without a structural approach. For example, the work can be improved with the inclusion of a theoretical model that aims to disentangle the channels of impact more robustly (especially those pertaining to political participation based on the political science literature), the impact and interaction of different immigration laws and the feedback loop between the ideological positions of candidates and voter behaviour.

[67]That said, I do study the 2014 elections to answer the same question in this paper and find consistent results as per the 2018 elections mainly in terms of a significant reduction in the electoral participation. I also report a significant increase in vote shares to centrist parties, an insignificant increase to right-wing and far-right parties and an insignificant decrease towards left-wing parties.

Acknowledgements I am thankful for the valuable suggestions by Professor Julien Labonne, members of the migration group at the Oxford Department of Economics, and members of the Quantitative Informal Workshop at the Oxford Department of International Development. I am also thankful to Dr. Barbara Souza for her valuable insights and discussion on the Brazilian political context. I am grateful to Leonardo Cavalcanti, Tadeu Oliveira, Marília Fernandes, and all the staff of the Observatório das Migrações Internacionais (OBMigra) at the Universidade de Brasília for their initial facilitation of the data on migration and for answering questions about it.

Appendices of Chapter 3

III.A Supplementary Material: Ideology Categorisations

In order to identify the political party on an ideology spectrum, I follow the steps that have been used in Rozo and Vargas (2021) which build on the detailed methodology employed by Fergusson et al. (2020). The datasets I use, link each candidate to their political party aligned with at the moment of candidacy. The steps I perform are outlined as follows:

- Identify the party names, slogans, mottos, and description. I then look out for key words that would help categorise the party as left, centre or right wing. An example of such words would be: "communist", "socialist", "conservative" etc.
- For parties where the description in the party logo or motto is unclear, I research the respective parties' or candidates positions on policy stances. Following Rozo and Vargas (2021), a party is considered to be a "left" wing party if this party holds 1- pro-peasant views, 2- more government intervention in the market, 3- defends workers' rights and protection, 4- advocates state-owned enterprises and 5- are anti-imperialist. In contrast, a party is deemed to be right-wing if a party stands for 1- favoristism towards growth promoting policies as opposed to redistribution, 2- free market oriented policies, 2- family and religion as key pillars of society, 4- believes in patriotism and nationalism, 5- favours a stronghold policy towards ensuring security, law and order.
- Parties and candidates that show a mix of both of these sets of categories are considered to be centrist parties.
- Depending on the extent of which parties agree with certain policies, I classify them as far left or far right.
- Finally, I complement and adjust this analysis based on the position of the candidate running for elections on different topics as specified by (Wiesehomeier et al., 2021) and according to the ideological classification by Baker and Greene (2019).

Table III.13 shows the ideological classification for all parties in Brazil, by the party acronym.

Table III.13: Brazilian Party Codes and Ideology Classifications

Party Code	Ideology Classification	Party Code	Ideology Classification
AN	centre	PC do B	far-left
Avante	centre	PCB	far-left
PAN	centre	PCML	far-left
PIRATAS	centre	PCO	far-left
PND	centre	PCR-UP	far-left
PRN	centre	PSPB	far-left
PROS	centre	PSTU	far-left
PSD	centre	APB	far-right
PSN	centre	PRTB	far-right
PV	centre	PMN	left-wing
SD	centre	PNC	left-wing
Solidariedade	centre	PPS	left-wing
PSB	Centre-centre-left	PSP	left-wing
REDE	Centre-centre-left	PST	left-wing
MDB	Centre-centre-right	PT	left-wing
PFL	Centre-centre-right	PTR	left-wing
PMDB	Centre-centre-right	UP	left-wing
ANIMAIS	centre-left	PSOL	left-wing far-left
Cidadania	centre-left	ARENA	right-wing
PGT	centre-left	DC	right-wing
PDT	centre-left left-wing	DEM	right-wing
PPL	centre-left left-wing	LIBER	right-wing
PS	centre-left left-wing	PACO	right-wing
PSDB	centre-right	PDS	right-wing
PCN	centre-right	PMB	right-wing
PH	centre-right	PODE	right-wing
PHS	centre-right	PP	right-wing
PL	centre-right	PPB	right-wing
PMC	centre-right	PSDC	right-wing
PR	centre-right	PTC	right-wing
PTB	centre-right	PTN	right-wing
PRP	centre-right right-wing	Republicanos	right-wing
NOVO	centre-right right-wing	UDN	right-wing
PRB	centre-right right-wing	PATRI	right-wing far-right
PT do B	centre-right right-wing	PDC	right-wing far-right
		PEN	right-wing far-right
		PRONA	right-wing far-right
		PSC	right-wing far-right
		PSL	right-wing far-right
		RDP	right-wing far-right

Note: This table shows the ideology classifications resulting from the exercise outlined in section III.A. The following party codes represent all parties in Brazil.

III.B Supplementary Material: IV Shift Share Robustness Checks

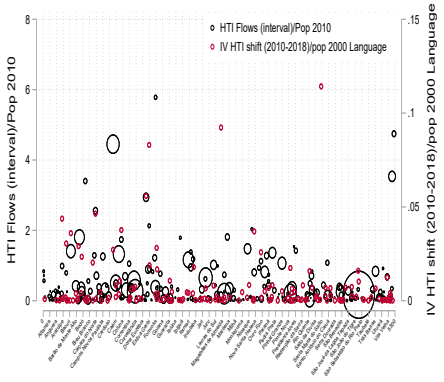
III.B.1 IV Shift Share Validity and Robustness Tests

The Shift Share instrumental design has been used extensively in the literature to obtain plausible exogenous variation in a variety of settings (Altonji and Card (1991), Card (2001), Borjas (2014), Card and Peri (2016), Aydemir and Kırdar (2014), Dustmann, et al. (2017), Monras (2020), Llull (2018), Angrist, et al. (1999), Anatolyev (2013), Chetty, et al. (2015)).

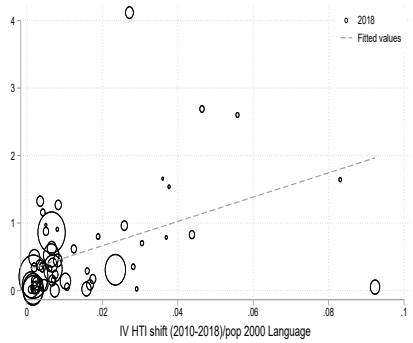
I present in the following figures the distribution of IV measure and the Haitian immigration to population measures as described and referred to in the text.

Figure III.11: Distribution of the Haitians and the calculated IV across municipalities

(a) Distribution of Haitian Immigration and IV



(b) Correlation between Immigration and IV



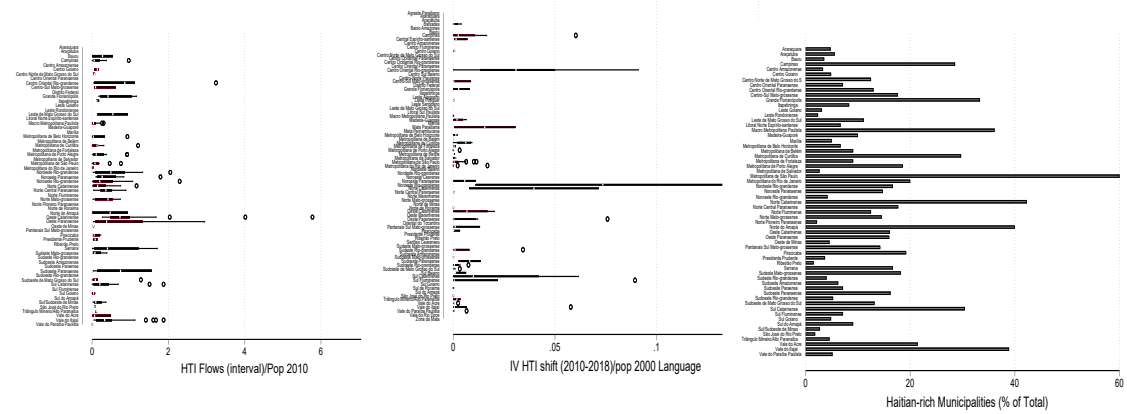
Notes: The figures show the distribution of the migrant share to population and the calculated IV Shift share. The Immigration to population variable is calculated as per equation (5.4) and the IV shown is as per the calculation in equation (5.6). Scatter plot is weighted by the absolute number of immigrants in a specific municipality.

Figure III.12: Distribution of the Haitians and the calculated IV within Mesoregions

(a) Distribution of Haitian Immigration

(b) Distribution of IV

(c) % of Municipalities with Haitian Immigrants by Mesoregion



Notes: The figures show the distribution of the migrant share to population within mesoregions. The Immigration to population variable is calculated as per equation (5.4). There are 137 mesoregions in Brazil, of which 62 have Haitian immigrants. I am excluding the Federal District from figure III.12(c) because it only has one municipality that received Haitian immigrants, making the ratio a 100%.

Since the IV shift share instrument combines a multitude of different instruments (the underlying shares), it is not always easy to assess the viability of the identifying assumptions. I attempt to uncover the black box of the instrument drawing on recent literature from Goldsmith-Pinkham et al. (2020) and Borusyak et al. (2018) to try and understand what the sources of variation within the instrument are.

Following the approach outlined by Goldsmith-Pinkham, Sorkin, and Swift (2020). The authors show that in finite samples, the IV 2SLS estimator is numerically equivalent to a generalised method of moments (GMM) estimator where the latter is a weighted sum of just-identified estimates based on a finite sample decomposition due to Rotemberg (1983). Through rethinking the Bartik instrument as a GMM estimator, the IV estimator can be decomposed into a set of estimators each representing the local shares and a set of Rotemberg weights associated with each instrument. Each of the just identified estimates with their associated Rotemberg Weights will provide the contribution of each share to the overall Bartik estimate. I refer readers to Goldsmith-Pinkham, Sorkin, and Swift (2020) for the proofs.

In this analysis, I exploit the quasi-natural experiment of the unexpected immigration shock from Haiti. However, I cannot rely on this exogeneity assumption because Haitians relocated non-randomly to different municipalities in Brazil. Therefore, to

identify the impact, I rely on many locations with historic settlements of immigrants who speak different languages prior to the year 2000. The assumption in this analysis is that the research design reflects differential exposure of municipalities to the common shock of Haitian flows. Therefore the exogeneity condition assumed in my study is that the shares of migrants who speak a certain language n at time $t = 0$ are exogeneous to voting outcomes in time $t = 1$. Goldsmith-Pinkham et al. (2020) then argue that is a researcher is to embrace this "shares" view of the identification of the instrumental variable strategy, then they must provide evidence that the municipalities with different initial stock of immigrants would have counterfactually evolved in a similar manner to other municipalities. In order to test the plausibility of the identifying assumption, I run three empirical tests as recommended by Goldsmith-Pinkham, Sorkin, and Swift (2020).

The first empirical test I perform are overidentification tests. Overidentification test present a formal test to misspecification and whether or not the instruments are correlated with the error terms beyond what would be expected by chance and relies on the validity of one of the instruments. I report the results across different estimators in table III.14. Since I am assuming heteroskedasticity, I rely on the Wooldridge robust score test to check for overidentification, and I find that I fail to reject the null which indicates that the model is correctly specified or that there is at least one valid instrument. Using the Bartik instrument as the summation of the shares, I find that the estimates are very close to one another between columns (4) to (6); i.e. across the 2 SLS, Limited Information Maximum Likelihood (LIML), and Modified Bias-corrected TSLS (MBTSLS).

Table III.14: Estimates on centrist vote shares

	(1) Centre Parties	(2) Centre Parties	(3) Centre Parties	(4) Centre Parties	(5) Centre Parties	(6) Centre Parties
HTI Flows (interval)/Pop 2010	0.595***	0.667***	-0.0750	0.781**	0.841***	0.788*
	(0.212)	(0.227)	(0.106)	(0.306)	(0.322)	(0.451)
Observations	5558	5558	5558	5558	5558	5558
Estimation	TSLS	LIML	OLS	TSLS (Bartik)	LIML (Bartik)	MBTSLS (Bartik)
p-value of overid test	.258	.129				
Municipality Controls	Yes	Yes	Yes	Yes	Yes	Yes
Mesoregion F.E	Yes	Yes	Yes	Yes	Yes	Yes

Note: This table shows the second stage estimates on the vote shares for the centrist party in the Presidential elections. I present a variety of estimates at the municipality level in 2018. In column (1) and (2), I use all the underlying language instruments as separate instruments to predict the first stage Haitian immigrant distribution, using 2 SLS and the Limited Information Maximum Likelihood (LIML). For these two estimators, I present the overidentification tests. For column (1), I use the Wooldridge robust score test, and since a robust option is not available for the LIML estimation in column (2), I present Sargan's χ^2 test. Columns (3) to (6) present the OLS, 2SLS, LIML, and the Modified Bias-corrected TSLS (MBTSLS), with the latter three using the Bartik IV as the instrument.

As a second step within this empirical analysis, I calculate the Rotemberg Weights with controls[68] for 2018. I find that in my analysis the distribution of sensitivity is skewed towards a few languages.[69] The top three instruments, Haitian, Spanish and English, account for 98% (0.994/1.01) of the positive weight in the estimator. These weights are important in reflecting the variation in the data that the Bartik IV is using. In this setting, one of the most important comparisons is across municipalities with different variation of Haitian migrants (who account for approximately 51% of the weight in the estimator).[70] Therefore my estimates will be sensitive to deviations from the identifying assumptions governing the spatial distribution of Haitian immigrants. Panel A summarises the weights α_f decomposed by positive and negative Rotemberg weights. One can see that the positive weights constitute the highest share of overall weight. Panel B explores the correlation between these shares, the national migrant

[68] I use poverty rate and GDP per capita in the base year of 2010.

[69] Goldsmith-Pinkham, Sorkin, and Swift (2020) find a similar results in their analysis of the work done by Autor, Dorn and Hanson (2013) on the impact of Chinese imports on manufacturing employment in the United States and of the work done by Card (2009) in order to estimate the negative inverse elasticity of substitution between immigrants and natives.

[70] This could be driven by also having a small number of the Haitian influx into the country, so the share from the total will also be relatively high.

flows, the just identified estimates β_f, the first stage F-test and the variance of the shares of this instrument z_f. The Rotemberg weights are positively correlated with the national migrant flows and the variance of the local language shares. Higher dispersion in the latter implies a higher variation in the cross sectional exposure to initial language group settlements, and hence creating higher variation in the instrument.

Upon unpacking the underlying shares driving the power of the instrumental variable in the first stage, I find that more than 50% of the variation is attributed to the Haitian original settlements in Brazil. This is interesting for two reasons. First, it indicates that in this context, it is the case that power of the instrument is driven by the original settlements of the country of origin, the power of the enclave effect of having a concentration of immigrants from the same country group in the same area. With this in mind, it is not then surprising that the earlier settlements of Haitians themselves are driving the power of the first stage in the instrumental variable approach. A further deeper investigation on the role of linguistic proximity, including experimenting with different linguistic proximity measures is a promising avenue for future research, but goes beyond the scope of this paper's purpose. This analysis would entail accounting for multiple languages spoken within the same country and probably excluding other countries that can share a language but are very different in other aspects (economic development, religion, etc.). The second reason this finding is interesting is because the results are driven by those earlier Haitian settlements themselves as opposed to, for example, French speakers that are more likely to be voluntary immigrants, from a different race to the Haitians, and come from a country with a higher level of economic development, and therefore may influence how natives react differently.

Table III.15: Rotemberg Weights and Contribution of Language Shares

Panel A: Negative and Positive Weights

	Sum	Mean	Share
Negative	-0.010	-0.001	0.010
Positive	1.010	0.029	0.990

Panel B: Correlations of Language Groups

	α_k	g_k	β_k	F_k	Var(z_k)
α_k	1				
g_k	0.210	1			
β_k	-0.074	-0.230	1		
F_k	0.047	0.007	-0.117	1	
Var(z_k)	0.811	0.059	0.097	-0.082	1

Panel C: Top 3 Rotemberg Weight Languages

	$\hat{\alpha}_k$	$\hat{\beta}_k$
Haitain	0.519	1.439
Spanish	0.369	1.947
English	0.106	3.553

Note: α_f represents the Rotemberg Weights calculated for every language; g_f represents the national flows of migrants who speak language f; β_f represents the contribution of each instrument to the overall Bartik instrument and z_f represents the iv for every language f.

The second empirical test I perform is I study whether the shift-share IV and the main language groups in terms of their contribution to the weight of the IV are plausibly exogenous to shocks affecting voting outcomes. I regress the Bartik IV and the language shares on observable characteristics for the year 2000 ($t = 0$) by estimating the following equation and report the results in table III.16:

$$Z_m = \alpha + \gamma_1 X_{m,0} + \theta_r + \epsilon_m. \tag{III.B.14}$$

where Z_m is the Bartik IV or the individual language shares of $\in$

$Haitian, English, Spanish$. Studying the correlation for the initial time period is a tighter restriction and reflects the instrument's cross sectional variation. If the shares at time $t = 0$ are correlated with potential confounding factors, then this might be indicative that there are other omitted variables biasing the estimation. As seen in table III.16 there is weak to no correlation between the shares and variables potentially correlated with voting. Across the three language shares and the Bartik instrument, there is a correlation with VA added services that is significant at the 5% level at least. I make sure to control for this variable among others in all my estimations. Also the R-squared ranges between 0.01 to 0.03 are consistently small suggesting that languages contributing to the most variation are not systemically located in certain municipalities in Brazil.

Table III.16: Correlation of Shares with Municipality Characteristics in 2010

	(1) IV HTI shift (2010-2018)/pop 2000 Language	(2) Haitian	(3) Spanish	(4) English	(5) IV HTI alone
Poverty Rate	0.00225**	1.26e-08*	6.15e-09*	2.93e-09	0.00166*
	(0.00102)	(6.82e-09)	(3.46e-09)	(2.09e-09)	(0.000951)
Unemployment Rate	-0.000317	-2.50e-09	2.67e-10	1.44e-09	-0.000212
	(0.00106)	(6.93e-09)	(4.31e-09)	(3.09e-09)	(0.000956)
Sanitation Coverage	0.00107	9.15e-09**	3.34e-09	3.64e-09	0.00118*
	(0.000688)	(4.39e-09)	(2.98e-09)	(2.39e-09)	(0.000608)
Percapita GDP	0.000551**	2.24e-09	1.75e-09*	7.29e-10**	0.000276
	(0.000266)	(1.42e-09)	(1.04e-09)	(3.00e-10)	(0.000194)
Employed in agriculture	-0.00106	-9.62e-09	1.35e-09	3.11e-09**	-0.00154
	(0.000937)	(6.71e-09)	(2.10e-09)	(1.41e-09)	(0.000934)
Water and Trash Coverage	0.00000280	-1.75e-09	1.73e-09	5.47e-10	-0.000293
	(0.000537)	(3.41e-09)	(1.46e-09)	(6.64e-10)	(0.000472)
Ratio of White/Brown income	0.000419	-5.11e-10	2.09e-09	-6.39e-10	-0.0000403
	(0.000729)	(5.04e-09)	(2.90e-09)	(1.74e-09)	(0.000691)
Ratio of White/Black income	-0.000234	5.51e-10	-3.49e-10	3.09e-09	-0.0000908
	(0.000703)	(4.95e-09)	(3.09e-09)	(2.27e-09)	(0.000676)
Ratio of Black/Brown income	0.000237	-1.46e-09	2.08e-09	-5.88e-10	-0.000160
	(0.000505)	(3.45e-09)	(2.21e-09)	(1.24e-09)	(0.000470)
Tax	0.00388	2.36e-08	-8.98e-10	8.39e-09*	0.00262
	(0.00400)	(2.03e-08)	(1.37e-08)	(4.35e-09)	(0.00277)
Employed	0.0140	9.06e-08	1.32e-08	1.76e-08	0.0125
	(0.0125)	(8.44e-08)	(2.22e-08)	(1.68e-08)	(0.0118)
Paid ppl	-0.00910	-6.71e-08	7.42e-09	-1.27e-08	-0.00941
	(0.0126)	(8.17e-08)	(3.16e-08)	(1.75e-08)	(0.0114)
Company	-0.0111	-8.46e-08	-1.16e-08	1.44e-08	-0.0115
	(0.0145)	(9.98e-08)	(3.26e-08)	(1.94e-08)	(0.0139)
VA agri	0.000885	5.66e-10	7.50e-09**	2.75e-09	-0.0000267
	(0.000820)	(4.97e-09)	(3.75e-09)	(1.92e-09)	(0.000665)
VA Services	0.00253***	1.43e-08***	7.33e-09***	4.30e-09**	0.00184***
	(0.000674)	(4.51e-09)	(2.53e-09)	(1.75e-09)	(0.000621)
Constant	-0.00732***	-2.65e-08**	-2.94e-08**	-1.59e-08***	-0.00296*
	(0.00263)	(1.23e-08)	(1.17e-08)	(5.65e-09)	(0.00166)
Observations	5543	5543	5543	5543	5543
MUN with Past HTI	245	50	173	52	50
Year	2010	2010	2010	2010	2010
Sample	All*	All*	All*	All*	All*
Weighted Regression	Yes	Yes	Yes	Yes	Yes
Adj. R sq	.035	.018	.039	.019	.015
Municipality Controls	Yes	Yes	Yes	Yes	Yes
Mesoregion F.E	Yes	Yes	Yes	Yes	Yes

Notes: The table shows the 2SLS first stage estimates. Robust standard errors in parentheses clustered at the mesoregion level. Regressions include mesoregion fixed effects and municipality controls in 2010. Results are shown for two sample: full sample of municipalities and municipalities that received at least one migrant. The predicted Haitian migration variable is adjusted by population in 2010. Data sources: IBGE 2010 Census data for control variables, Tribunal Superior Eleitoral and SISMIGRA for migration data.

The third empirical test I perform is to check if historical language migrant shares are correlated with pre-trends in the outcomes leading up to the migration, which first started in 2010. This is also important because in the construction of the IV, I rely on early settlement of immigrants that, by the time of the study, 2018, might have become naturalized migrants and hence with the ability to vote. To do so, I estimate the following the equation and report the results in figure III.13:

$$E_{m,t} = \sum_{T=1998(4)2006} [\sigma_T(L_m) \times (t = T)] + \theta_{r,t} + \epsilon_{m,t}, \tag{III.B.15}$$

where $\theta_{r,t}$ is mesoregion-year fixed effects[71]. I find that there is no correlation between the share component of the Bartik instrument and the political outcomes (vote share to the left-wing parties, centrist parties, and turnout ratio) in the years leading up to the Haitian immigration shock. I investigate further the pre-outcomes on turnout ratio and whether earlier trends in this variable were a predictor to immigrant locations. I report the results of this analysis in table III.17. In column (1), I regress the turnout ratio on the language share interacted by time trend from 2002 to 2018, omitting the year of the immigration onset, 2010. I find that there was positive correlation in the year 2002 between the language shares and the turnout ratio, but that significance dissipated in later years (2006) and then turned negative in the years after the immigration shock (2014 and 2018). In column (2), I present the correlation between the language share and the turnout ratio for all the years leading up to the Haitian immigration shock, and find a non-significant correlation. Column (3) estimates a similar relation to column (2) but adding mesoregion- year fixed effects and municipality level controls. Columns (4) to (6) ask if the earlier municipality level turnout ratios predicted the new influx of Haitian immigration and I find that there is no significant correlation.

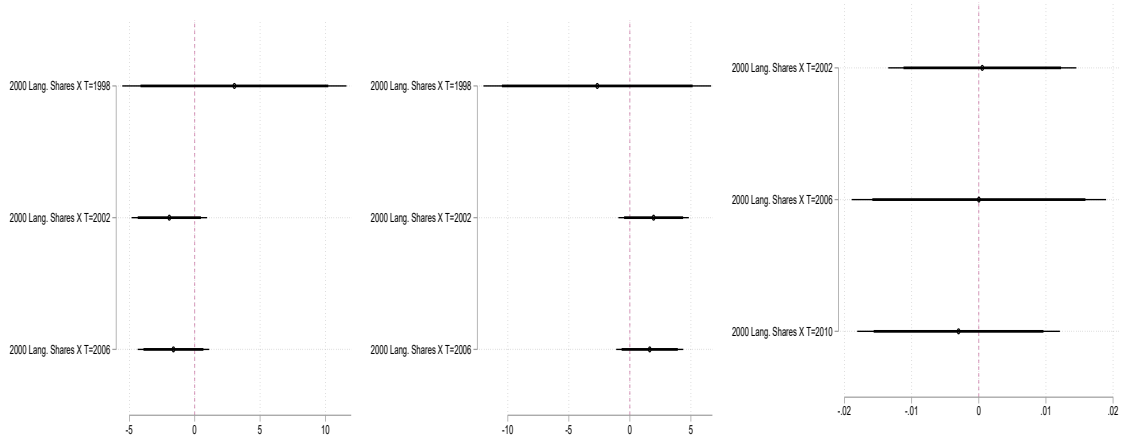
[71] I try this same estimation with mesoregion and year fixed effects not interacted and I actually find that the share interacted with 2006 has no impact on turnover while earlier shares interacted with 1998 and 2002 respectively have a positive impact on turnover ratio. Similarly for the impact on votes to the centre, there is no impact by there is a negative impact through earlier shares that is positive.

Figure III.13: Testing for Pre-trend Correlations

(a) Pre-Trends in the Presidential Left-wing Party votes

(b) Pre-Trends in the Presidential centrist Party votes

(c) Pre-Trends in the Turnout



Notes: For the turnout ratio, it is imputed as the number of votes registered for all parties as a proportion of all voters.

Table III.17: Language Share impact on Pre-trends

	(1)	(2)	(3)	(4)	(5)	(6)
	Turnout Ratio (%)	Turnout Ratio (%)	Turnout Ratio (%)	HTI Flows (interval)/Pop 2010	HTI Flows (interval)/Pop 2010	HTI Flows (interval)/Pop 2010
2000 Lang. Shares X T=2002	2.463***					
	(0.523)					
2000 Lang. Shares X T=2006	1.272					
	(0.989)					
2000 Lang. Shares X T=2014	-0.666					
	(1.166)					
2000 Lang. Shares X T=2018	-3.060					
	(1.902)					
2000 Lang. Shares		-0.0819	0.396			
		(0.805)	(1.145)			
Turnout Ratio (%) 2002				-0.0000999		
				(0.000514)		
Turnout Ratio (%) 2006					-0.000485	
					(0.000545)	
Turnout Ratio (%) 2010						-0.000994
						(0.000612)
Observations	27775	16680	16656	5550	5552	5554
MUN with Past HTI						
Year	All −2010	2002-2010	2002-2010	2002	2006	2010
Sample	All*	All*	All*	All*	All*	All*
Weighted Regression	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R sq	.5365289485230029	.4706465195391722	.5989361974091602			
Municipality Controls	No	No	Yes	Yes	Yes	Yes
Mesoregion F.E	Yes (Meso Year)	Yes	Yes (Meso X Year)	Yes (Meso Year)	Yes (Meso Year)	Yes (Meso Year)

Notes: This shows the pre-trends of the shares Data sources: IBGE 2010 Census data for control variables, Tribunal Superior Eleitoral and SISMIGRA for migration data.

In what follows I attempt to assuage the worry that characteristics of municipalities that attracted immigrants pre-2000 could have persistent and confounding impacts that could influence Haitian migration decisions from 2010 onwards and on political behaviour of natives in 2018. In table III.18 and III.19, I show that the pre-period change in outcomes, so the difference between 2010 to 2000, are uncorrelated with the subsequent predicted immigration by the instrument. This evidence helps rule out concerns of reverse causality.

Table III.18: Pre-Trends Correlations on Predictor Variables

"Dep. Variable is 2010-2000 change"	(1) Turnout Ratio	(3) Agri ppl	(4) Average Salary	(5) Mun Revenue	(7) Mun Exp	(8) Per Capita GDP	(9) Poverty Rate	(11) Unemployment
HTI Flows (interval)/Pop 2010	0.0969	0.0249	105.6	758985558.0	-4921196.5	-1501.5	0.00852	-0.00861
	(0.412)	(0.0730)	(209.8)	(614443755.9)	(8962962.2)	(1016.1)	(0.00655)	(0.00858)
Observations	5550	2762	2700	2569	2550	2753	2798	2705
Sample	All*	All*	All*	All*	All*	All*	All*	All*
Weighted Regression	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Municipality Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mesoregion F.E	Yes (Meso Year)	Yes (Meso Year)	Yes (Meso Year)	Yes (Meso Year)	Yes (Meso Year)	Yes (Meso Year)	Yes (Meso Year)	Yes (Meso Year)

Notes: This table shows the second stage results of a 2SLS estimation. The outcome variables are the pre-period change (2010-2000) of the variable, regressed on instrumented immigration. Data sources: IBGE 2010 Census data for control variables, Tribunal Superior Eleitoral and SISMIGRA for migration data.

Table III.19: Pre-Trends Correlations on Outcome Variables

"Dep. Variable is 2010-2000 change"	(1) Presidential Left-wing	(2) Presidential Right-wing	(3) Presidential Centre-wing	(4) Governor Left-wing	(5) Governor Right-wing	(6) Governor Centre-wing
HTI Flows (interval)/Pop 2010	0.651	0.0150	-0.665	-0.692	2.196	-1.504
	(1.630)	(0.0848)	(1.650)	(1.642)	(1.399)	(1.846)
Observations	5551	5551	5551	5551	5551	5551
Sample	All*	All*	All*	All*	All*	All*
Weighted Regression	Yes	Yes	Yes	Yes	Yes	Yes
Municipality Controls	Yes	Yes	Yes	Yes	Yes	Yes
Mesoregion F.E	Yes (Meso Year)	Yes (Meso Year)	Yes (Meso Year)	Yes (Meso Year)	Yes (Meso Year)	Yes (Meso Year)

Notes: This table shows the second stage results of a 2SLS estimation. The outcome variables are the pre-period change (2010-2000) of the variable, regressed on instrumented immigration. Data sources: Tribunal Superior Eleitoral and SISMIGRA for migration data.

In addition to the empirical tests mentioned, I perform an additional test to check if my shift-share IV is conflating labor demand factors or other local pull factors that could influence the location choices of other immigrants (besides Haitians). I report the results in table III.20. I find that the IV based on language proximity to Haitian Creole does not predict the location of migrants who are not from Haiti, across different visa categories.

Table III.20: IV Predicting Non-Haitian Immigrants

	(1)	(2)	(3)	(4)
	Other Asylum	Other Permanent	Other Temporary	Other Resident
IV First Stage: Haitian Migration				
IV HTI shift (2010-2018)/pop 2000 Language	1.248	2.779	7.040*	6.475
	(0.930)	(1.802)	(4.274)	(5.946)
Observations	5558	5558	5558	5558
Year	2018	2018	2018	2018
Sample	All	All	All	All
Kleibergen-Paap F-stat	1.8001	2.3798	2.7134	1.1859
Anderson-Rubin Test	0.4479	0.4479	0.4479	0.4479
Cragg-Donald F Stat	72.8038	61.9082	35.2601	52.0923
Municipality Controls	Yes	Yes	Yes	Yes
Mesoregion F.E	Yes	Yes	Yes	Yes

Notes: This table shows the first stage results of a 2SLS estimation using the IV based on the Haitian Creole language. The outcome variables are the ratio of other immigrants that hold different types of visas to the population. Data sources: IBGE 2010 Census data for control variables, Tribunal Superior Eleitoral and SISMIGRA for migration data.

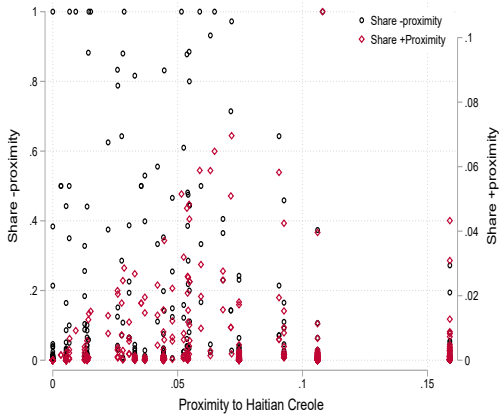
The language proximity index can be thought of as a weighting assigned to a particular language group that existed pre-2000. For a constant ratio of $\frac{Mig_{f,m}}{Mig_f}$ in 2000, each language group receives a different weight, through λ_f, in predicting the locations of Haitians. The importance of this language proximity index is that it transforms the shares of different language groups into a *Haitian equivalent migrant/ratio* that settled into Brazil prior to 2000. One can imagine a scenario where two historic language groups, one that is proximate to Haitian Creole with one that is very distant, but the former has a much smaller $\frac{Mig_{f,m}}{Mig_f}$. In this case, the interpretation of the language share L_m remains the same: it represents the Haitian equivalent ratio in a certain municipality, because the language with the larger concentration share is being "penalised" by it's linguistic distance from Haitian Creole, while the other is being accentuated.

I visualise in figure III.14 how the language proximity index penalises the ratio of the language shares. This graph depicts the share component without the language proximity index and the share component with the language proximity index against

λ, the language proximity index. The graph shows that the shares of language groups are transformed depending on the value of the language proximity index for each language group. In essence, each language group is weighted differently depending on the proximity of this language to Haitian creole. Naturally, the highest weight is given to the Haitian Creole speakers themselves, the Haitians.

What is key in the construction of this instrument is that the share component, L_m, represents the share of immigrants from a certain language group who reside in municipality m as a proportion of all immigrants from language f who came into Brazil during this period $Mig_{f,2000}$. This construction is similar to other shift-share IV strategies relying on the past settlement instrument, where the denominator is the overall total number of people from a certain country (Mayda et al., 2022). This ratio does highlight the concentration of immigrants in a municipality taking into account the overall migration volume of immigrants in this language group. For a given number of immigrants per municipality, my IV is permitting municipalities that receive immigrants from a small language immigration wave to have more predictive power than larger language immigrant waves. I have experimented with a different form of the IV where the denominator is the total number of immigrants per municipality ($Mig_{m,2000}$). The difference between having the denominator as $Mig_{f,2000}$ versus $Mig_{m,2000}$ is that the former is saying, for example, how many Spanish immigrants chose to locate in São Paulo from all the Spanish speakers that came into Brazil, as opposed to saying how many Spanish people located in São Paulo as a proportion of all immigrants in São Paulo, in the latter. In cities with a lot of immigrants, more than those that came from Spanish speaking countries, can dilute the presence of Spanish immigrants ($Mig_{m,2000} > Mig_{spanish}$) and hence change their relative importance in predicting the location of new Haitians immigrants. In cities, where there is not a lot of immigrants, and Spanish speakers located there historically, ($Mig_{m,2000} < Mig_{spanish}$), having the denominator $Mig_{m,2000}$ in the share will overemphasise the importance of this location to Spanish speakers, while in reality a small proportion from those immigrants who came into Brazil located there. I confirm after calculating both IV strategies and estimating the first stage equation that having the denominator as $Mig_{m,2000}$ predicts the locations of new immigrant waves much less than having the denominator as $Mig_{f,2000}$.

Figure III.14: Discussion on the IV Share Component



This graph depicts the share component without the language proximity index and the share component with the language proximity index against λ, the language proximity index.

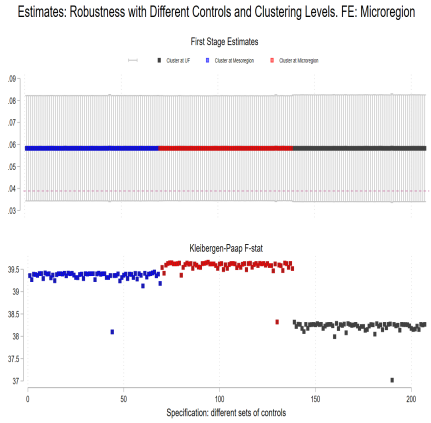
Table III.21: IV Leave one-out strategy at the Microregion

	(1)	(2)	(3)	(4)
	HTI Flows (interval)/Pop 2010	HTI Flows (interval)/Pop 2010	HTI Flows (interval)/Pop 2010	HTI Flows (interval)/Pop 2010
IV First Stage: Haitian Migration				
iv_htin_p_r_mi	20.17***	18.47***	20.23***	17.88***
	(3.843)	(3.110)	(3.784)	(3.204)
Observations	5558	316	5558	316
Year	2018	2018	2018	2018
Sample	All	Nonzero	All	Nonzero
Weighted Regression	Population Weighted	Population Weighted	Nonweighted	Nonweighted
Kleibergen-Paap F-stat	27.5398	35.2767	28.5916	31.1344
Anderson-Rubin Test	0.4200	0.6054	0.2863	0.6141
Cragg-Donald F Stat	867.7316	34.0069	809.4472	30.1956
Municipality Controls	Yes	Yes	Yes	Yes
Mesoregion F.E	Yes	Yes	Yes	Yes

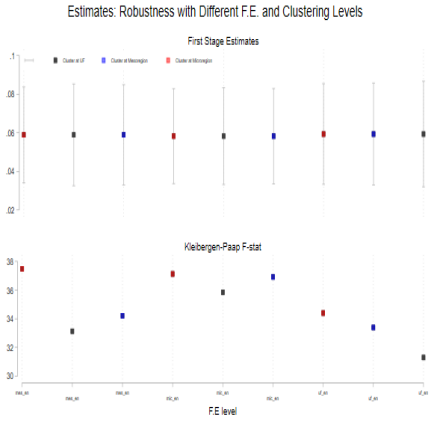
Notes: This table shows the first stage results of a 2SLS estimation using the IV based on the Haitian Creole language. The IV constructed with a leave one out strategy, where by instead of subtracting out the migrants from own municipality from the total flows, I am subtracting the total number of Haitian immigrants that settled in that microregion instead. Data sources: IBGE 2010 Census data for control variables, Tribunal Superior Eleitoral and SISMIGRA for migration data.

Figure III.15: Robustness Checks: Power of IV under Different F.E and Levels of Clustering

(a) Different Combinations of Control Variables at different levels of FE and Clustering of Standard Errors



(b) At different levels of FE and Clustering of Standard Errors



Notes: The graphs show the first stage estimates as well as the K-P F-statistic across different combinations of fixed effects levels, levels of clustering the standard errors and different combinations of the control variables. Three levels of fixed effects and clustering are considered: The microregion level (550) which is a level up from the municipality level studied, the mesoregion (136) which is a level up from the microregion and federative unit level (27) which is a level up from the mesoregion. The same levels for clustering are considered. Approximately 200 different combinations of control variables are considered.

III.B.2 Traditional Instrumental Variable Strategy

The traditional IV relies only on the earlier distribution on Haitian immigrants prior to the year 2000. In particular, the instrument is constructed as follows:

$$L_{hti,m} = (\frac{Mig_{HTI,m}}{Mig_{HTI}})_{t=0} \tag{III.B.16}$$

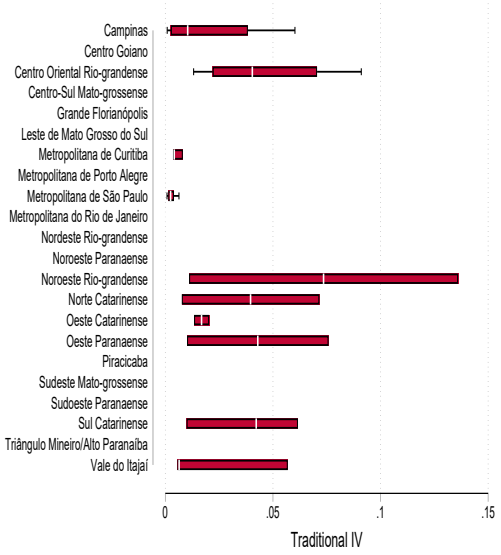
Where $Mig_{HTI,m}$ is the number of Haitians in municipality m at time 0 ($t_0 = 2000$). In order to get the instrument $Z_{hti,m,2018}$ for each municipality:

$$Z_{hti,m} = \frac{(L_{hti,m} * \sum_{2010}^{2018} Mig_{HTI,t}^{-m})}{\widehat{P_{m,2010}}} \tag{III.B.17}$$

Earlier Haitian settlements were present in 48 municipalities only, with São Paulo absorbing approximately 18% of the immigrants. I rely on this instrumental variable construction in order to estimate the results and discuss how they compare to the language proximity instrumental variable strategy. I portray the variation within the

mesoregion for this instrument in figure III.16.

Figure III.16: Distribution of Traditional IV



This graph shows the distribution of the traditional IV shift share instrument that relies on the Haitian immigration previous settlements within each mesoregion.

Table III.22: IV First Stage with Traditional IV

	(1) HTI Flows (interval)/Pop 2010	(2) HTI Flows (interval)/Pop 2010	(3) HTI Flows (interval)/Pop 2010	(4) HTI Flows (interval)/Pop 2010
IV First Stage: Haitian Migration				
IV HTI alone/pop	6.298***	8.007***	5.996***	7.365***
	(1.836)	(2.886)	(1.650)	(2.441)
Observations	5561	319	5561	319
MUN with Past HTI	48	48	48	48
Year	2018	2018	2018	2018
Sample	All	Nonzero	All	Nonzero
Weighted Regression	Population Weighted	Population Weighted	Nonweighted	Nonweighted
Kleibergen-Paap F-stat	11.7682	7.6995	13.1983	9.1023
Anderson-Rubin Test	0.1082	0.6764	0.1646	0.5654
Cragg-Donald F Stat	539.7623	18.9533	510.9230	16.5941
Municipality Controls	Yes	Yes	Yes	Yes
Mesoregion F.E	Yes	Yes	Yes	Yes

	(1) HTI Flows (interval)/Pop 2010	(2) HTI Flows (interval)/Pop 2010	(3) HTI Flows (interval)/Pop 2010	(4) HTI Flows (interval)/Pop 2010
IV First Stage: Haitian Migration truncating 99th IV percentile				
IV HTI alone/pop	9.092**	8.007***	8.123**	7.365***
	(4.416)	(2.882)	(3.873)	(2.438)
Observations	5560	318	5560	318
MUN with Past HTI	47	47	47	47
Year	2018	2018	2018	2018
Sample	All	Nonzero	All	Nonzero
Weighted Regression	Population Weighted	Population Weighted	Nonweighted	Nonweighted
Kleibergen-Paap F-stat	4.2393	7.7218	4.3995	9.1286
Anderson-Rubin Test	0.5936	0.6760	0.6579	0.5649
Cragg-Donald F Stat	526.5986	18.9533	469.5864	16.5941
Municipality Controls	Yes	Yes	Yes	Yes
Mesoregion F.E	Yes	Yes	Yes	Yes

Notes: The table shows the 2SLS first stage estimates. Robust standard errors in parentheses clustered at the mesoregion level. Regressions include mesoregion fixed effects and municipality controls in 2010. Results are shown for two sample: full sample of municipalities and municipalities that received at least one migrant. The predicted Haitian migration variable is adjusted by population in 2010. Data sources: IBGE 2010 Census data for control variables, Tribunal Superior Eleitoral and SISMIGRA for migration data.

Table III.23: IV First Stage with the Language Proximity Index using All pre-2000 Immigrants

	(1) HTI Flows (interval)/Pop 2010	(2) HTI Flows (interval)/Pop 2010	(3) HTI Flows (interval)/Pop 2010	(4) HTI Flows (interval)/Pop 2010
IV First Stage: Haitian Migration				
IV HTI shift (2010-2018)/pop 2000 Language	5.548***	6.074**	5.443***	5.833***
	(1.398)	(2.512)	(1.313)	(2.229)
Observations	5561	319	5561	319
MUN with Past HTI	243	243	243	243
Year	2018	2018	2018	2018
Sample	All	Nonzero	All	Nonzero
Weighted Regression	Population Weighted	Population Weighted	Nonweighted	Nonweighted
Kleibergen-Paap F-stat	15.7548	5.8493	17.1886	6.8470
Anderson-Rubin Test	0.7614	0.7549	0.6233	0.6265
Cragg-Donald F Stat	464.7818	11.7525	455.7457	10.9320
Municipality Controls	Yes	Yes	Yes	Yes
Mesoregion F.E	Yes	Yes	Yes	Yes

	(1) HTI Flows (interval)/Pop 2010	(2) HTI Flows (interval)/Pop 2010	(3) HTI Flows (interval)/Pop 2010	(4) HTI Flows (interval)/Pop 2010
IV First Stage: Haitian Migration truncating 99th IV percentile				
IV HTI shift (2010-2018)/pop 2000 Language	19.99***	18.22***	20.06***	17.65***
	(3.745)	(3.033)	(3.689)	(3.124)
Observations	5558	316	5558	316
MUN with Past HTI	240	240	240	240
Year	2018	2018	2018	2018
Sample	All	Nonzero	All	Nonzero
Weighted Regression	Population Weighted	Population Weighted	Nonweighted	Nonweighted
Kleibergen-Paap F-stat	28.4982	36.0900	29.5796	31.9314
Anderson-Rubin Test	0.4479	0.6181	0.3098	0.6242
Cragg-Donald F Stat	873.8626	34.0698	815.0831	30.2709
Municipality Controls	Yes	Yes	Yes	Yes
Mesoregion F.E	Yes	Yes	Yes	Yes

Notes: The table shows the 2SLS first stage estimates. Robust standard errors in parentheses clustered at the mesoregion level. Regressions include mesoregion fixed effects and municipality controls in 2010. Results are shown for two sample: full sample of municipalities and municipalities that received at least one migrant. The predicted Haitian migration variable is adjusted by population in 2010. Data sources: IBGE 2010 Census data for control variables, Tribunal Superior Eleitoral and SISMIGRA for migration data.

As shown in table III.22, the first stage Kleibergen-Paap F statistics falls below 10 in a series of classifications especially when the outliers of the IV at the 99th percentile level are truncated. That said, the traditional approach relying on the country of origin produces results in first stage and second stage estimations that are similar qualitatively to those obtained using the IV used in this paper. This is not surprising because my IV builds on the traditional approach which is technically embedded in it and this result is consistent with the finding in this section that Haitian Creole has the highest weight in the variation of the IV as per Goldsmith-Pinkham et al. (2020). The fact that I find that the traditional approach remains valid and produces similar results is lending credit to the hypothesis that immigrant enclaves from the same country of origin are important predictors of future immigrant inflows. Therefore, because the language proximity IV used in my analysis, first, allows me to exploit more of the underlying variation in the historic spatial locations of immigrants, and, second, has a first stage Kleibergen-Paap F statistic for the different specifications studied, I chose to work with it in this paper. It is also an opportunity to experiment with a different way of thinking about how immigrants locate, beyond their country of origin to incorporate cultural proximity. However, I expect that before any claim is made about the superiority of

either IV strategy (which is not the purpose of this paper), more robustness checks and comparison tests should be done and across a variety of contexts.

III.C Supplementary Material: Robustness Checks related to Data Treatment

The STI-MAR Asylum dataset contains data on immigrants with no recorded year of entry, municipality of residence or entry. In the case of Haitians, as described earlier, those were a majority with an inactive status (approximately 90%). Further, qualitative evidence on the Haitian legalisation process, I assumed that those Haitians were later given a permanent or temporary visa if they stayed in Brazil. Also as per the OBMigra annual report, immigrants with inactive status were dropped from the sample. However, I note that there are other immigrants in the sample (approximately 44,000) that also have no records on their year of entry or residence. Those immigrants if they have entered before the year 2000, would have been included in the construction of my instrument and so will only impact my results through their power and relevance in the first stage 2SLS estimation. Unfortunately, because of the absence of information on their municipalities of entry or residence, I have no choice but to drop them from the sample with the assumption that those are later captured by SISMIGRA dataset that registered the immigrants under a different visa status. The majority of these immigrants are males (79.5%) born between 1987 and 1977. Their marital status and racial composition is missing. In SISMIGRA, there are a lot of missing recorded municipalities of residence. For those that are relevant for my analysis, are the immigrants that arrived to Brazil before 2000 and the Haitians that arrived after 2010. I find that approximately 22,000 immigrants who arrived before 2000 had no recorded place of residence. I perform a balance test on observable characteristics between immigrants that arrived before 2000 and those that arrived before 2000 but have no residential information. If those immigrants with missing information were fundamentally different then they could have also chosen different locations. When it comes to the Haitians, among the earlier settlements (before 2000) there were only two Haitians and 849 (between 2010-2018) with an unknown location. These are limitations in the dataset

chosen that I have to face[72]. I perform a balance test on observable characteristics to see how different those immigrants are, and hence if that could be suggestive of different historical location decisions. I report the results in tables III.26 and III.27. Since the sample size will play a big difference in the t-test difference in means, I also report the normalised differences as per the recommendations by Imbens and Rubin (2007), and deem those as sufficiently different if the normalised difference exceeds one quarter (Imbens and Wooldridge, 2009). In the SISMIGRA dataset, I do not find that the normalised difference exceeds 0.25 on any of the individual characteristics. However, for the STI-MAR dataset, I find that there are clear large normalised differences between immigrants that have a recorded year of entry or registration versus those that didn't. In particular, immigrants with a recorded municipality had less males, less individuals within the age group of 18-39, and had more active status.

In cases where the municipality of residence is missing, I equated that with the municipality of entry of the immigrant, or a missing year of entry where I equate it with the year of registration. This is a limitation of the analysis. I perform multiple robustness checks by dropping prominent municipalities of entry and I find that my first stage K-P F stat is still strong. Majority in my sample have an active status and I do not drop those with an inactive or different status in this analysis if they have a declared municipality of residence or entry[73]. The main municipalities of entry for Haitians that are recorded in the dataset are Bonfim, Boa Vista, Corumba, Foz do Iguacu, Pacaraima, Guraulhos, Epitaciolandia, and Sao Pãulo. Further there are other municipalities that are important entry points for other immigrants, and maybe less so for Haitians. I also exclude all those municipalities that had an entry of more than a 1000 migrants and test the first stage of the instrument. I find that the first stage results are robust to inclusion or exclusion of such municipalities. I note that some of these municipalities (Pacaraima, Boa Vista, Bonfim, Macaé) have outlier IV of more than 99% of the sample, and I have excluded them from the main analysis of this paper. I report the results of these exercises in table III.24.

[72]For the SISMIGRA dataset, this number is 22,206 immigrants. However, for the Asylum seekers dataset, this number is much higher and reaches approximately 88,000 immigrants without a recorded location. This is unfortunately one of the limitations of this data. Future work that aims to publish this chapter in a peer reviewed journal will contact the Ministry of Economy and the relevant agencies in Brazil to check if there has been amendments to this limitation and to cross check this data source with other available datasets.

[73]If an immigrant has a declared municipality of entry or residence, they will remain informative for my analysis in constructing the past settlement shift-share IV.

Table III.24: First Stage Results: Haitian Migration Robustness

	(1)	(2)	(3)
	No Main Entry Mun for HTI	No Main Entry for All	No Asylum
IV First Stage: Haitian Migration- Sample Robustness			
IV HTI shift (2010-2018)/pop 2000 Language	20.09***	20.15***	18.88***
	(3.761)	(3.793)	(3.490)
Observations	5553	5552	5558
Year	2018	2018	2018
Sample	All	All	All
Weighted Regression	Yes	Yes	Yes
Kleibergen-Paap F-stat	28.5266	28.2074	29.2739
Cragg-Donald F Stat	880.2584	883.4976	868.9815
Municipality Controls	Yes	Yes	Yes
Mesoregion F.E	Yes	Yes	Yes

Notes: The table shows the 2SLS first stage estimates. Robust standard errors in parentheses clustered at the mesoregion level. Regressions include mesoregion fixed effects and municipality controls in 2010. Results are shown for two sample: full sample of municipalities and municipalities that received at least one migrant. The predicted Haitian migration variable is adjusted by population in 2010. This excludes the top 1 percentile of municipalities with a very high historic concentration. Those are two municipalities Macae and Poco das antes. Data sources: IBGE 2010 Census data for control variables, Tribunal Superior Eleitoral and SISMIGRA for migration data.

I perform further robustness checks on the data to check the first stage power of the IV and its sensitivity to the data sample. In particular, I do the following robustness checks and report the results in table III.25:

- I rely only on the sample of SISMIGRA without the earlier settlements of Asylum seekers. An asylum seeker could have entered Brazil on an Asylum visa in the earlier years before the year 2000, could have received another visa, such as a temporary or permanent one later, and is then recorded in the SISMIGRA dataset under a different visa. This could be a possibility and my dataset doesn't allow me to which of those immigrants changed status later on.[74]. I still believe

[74]The annual report of ObMigra (2020) confirms that it is not possible to apply duplicate treatment data across the datasets because data on asylum claims are obtained from from extracting part of the International Traffic System database – Alert Measures and Active Restrictions (STIMAR), under the responsibility of the Federal Police. As the extract obtained by OBMigra used does not have identifying variables of the migrant, it was not possible to apply any duplicate cleaning treatment.

that including Asylum seekers with their municipality of residence prior to 2000 is informative for the IV predictive power and reduces the chance that I am not capturing undocumented immigrants.[75] The location choice remains a relevant factor in informing future locations of new immigrants, under the exclusion assumption I make. In order to verify that I am not double counting earlier immigrants in the same location and thus giving the location more prominence, I rely only on SISMIGRA earlier settlements to predict current immigrant locations. I report this robustness check, also excluding the top 1 percentile of IV and the first stage remains strong as evident by the K-P F statistic.

- I exclude from the new waves of Haitian immigrants, between 2010 and 2018, those with an Asylum visa. I find that the IV remains strong.

- I test the sensitivity of the IV to the combination of controls used and the level of fixed effects used and the level of clustering used. I report the results in figure III.15(a). I find that the first stage estimates remain within a constant range and the K-P F statistic oscillates between 30 to 60.

[75]Based on the reported year of entry, there are STI-MAR (in asylum data) countries of immigration that are not present in SISMIGRA before 2000. There are 12 countries (Guinea; Ethiopia; Saudi Arabia; Somalia; Afghanistan; Albania; Dominica; Ghana; Jordan; Sri Lanka; Nepal; Togo), with a few (from 1-4 migrants). I compare the countries that report similar years of entry across the datasets, and I find that the year of registration reported is different. Also for presence of immigrants that had a large presence such as Angola, the immigrants report different years of registration, which is indicative that the datasets are recording different immigrants.

Table III.25: First Stage Results: Haitian Migration Without Asylum Seekers

	(1)	(2)	(3)	(4)
	HTI Flows/Pop 2010 No Asylum	HTI Flows/Pop 2010 No Asylum	HTI Flows/Pop 2010 No Asylum	HTI Flows/Pop 2010 No Asylum
IV First Stage: Haitian Migration				
IV HTI no Asylum	9.855***	9.386***	9.851***	9.140***
	(1.940)	(1.793)	(1.919)	(1.790)
Observations	5556	312	5556	312
MUN with Past HTI	230	230	230	230
Year	2018	2018	2018	2018
Sample	All	Nonzero	All	Nonzero
Weighted Regression	Population Weighted	Population Weighted	Nonweighted	Nonweighted
Kleibergen-Paap F-stat	25.8040	27.3970	26.3410	26.0839
Anderson-Rubin Test	0.0811	0.9067	0.0842	0.9901
Cragg-Donald F Stat	572.2524	27.3120	538.4869	25.1879
Municipality Controls	Yes	Yes	Yes	Yes
Mesoregion F.E	Yes	Yes	Yes	Yes

Notes: The table shows the 2SLS first stage estimates. Robust standard errors in parentheses clustered at the mesoregion level. Regressions include mesoregion fixed effects and municipality controls in 2010. Results are shown for two sample: full sample of municipalities and municipalities that received at least one migrant. The predicted Haitian migration variable is adjusted by population in 2010. This excludes the top 1 percentile of municipalities with a very high historic concentration. Those are two municipalities Macae and Poco das Antes. Data sources: IBGE 2010 Census data for control variables, Tribunal Superior Eleitoral and SISMIGRA for migration data.

Table III.26: Balance test on SISMIGRA dataset: Immigrants with Missing versus Not Missing Location

	(1)		(2)		T-test	Normalized
	Recorded MUN		No Recorded MUN		P-value	difference
Variable	N	Mean/SE	N	Mean/SE	(1)-(2)	(1)-(2)
Female	44720	0.396	22206	0.388	0.885	0.016
		(0.002)		(0.003)		
Male	44720	0.604	22206	0.612	0.885	-0.016
		(0.002)		(0.003)		
0 - 15 age	44720	0.212	22206	0.178	0.000***	0.084
		(0.002)		(0.003)		
15 - 25 age	44720	0.281	22206	0.252	0.000***	0.067
		(0.002)		(0.003)		
25 - 40 age	44720	0.357	22206	0.403	0.000***	-0.095
		(0.002)		(0.003)		
40 - 65 age	44720	0.139	22206	0.154	0.100	-0.043
		(0.002)		(0.002)		
65 - age	44720	0.008	22206	0.012	0.000***	-0.036
		(0.000)		(0.001)		
Missing	44720	0.002	22206	0.001	0.000***	0.024
		(0.000)		(0.000)		
Consensusally Separate	44720	0.000	22206	0.000	0.317	-0.012
		(0.000)		(0.000)		
Divorced	44720	0.001	22206	0.001	0.996	-0.001
		(0.000)		(0.000)		
Judicially Separate	44720	0.000	22206	0.000	0.190	0.008
		(0.000)		(0.000)		
Married	44720	0.451	22206	0.513	0.000***	-0.124
		(0.002)		(0.003)		
Missing	44720	0.000	22206	0.001	0.669	-0.007
		(0.000)		(0.000)		
Other	44720	0.069	22206	0.086	0.000***	-0.065
		(0.001)		(0.002)		
Seperated	44720	0.000	22206	0.000	N/A	N/A
		(0.000)		(0.000)		
Single	44720	0.442	22206	0.357	0.000***	0.172
		(0.002)		(0.003)		
Widower	44720	0.036	22206	0.041	0.006**	-0.029
		(0.001)		(0.001)		
Active	1256	0.991	328	0.979	0.193	0.119
		(0.003)		(0.008)		
Canceled	1256	0.002	328	0.012	0.115	-0.173
		(0.001)		(0.006)		
Error	1256	0.002	328	0.000	0.163	0.055
		(0.001)		(0.000)		
Excluded	1256	0.002	328	0.006	0.424	-0.066
		(0.001)		(0.004)		
Expired Deadline	1256	0.002	328	0.003	0.712	-0.013
		(0.001)		(0.003)		

Notes: This table presents the balance test between observable characteristics of immigrants with reported location of residence and those without. The table reports the T-test p-value and the normalised difference. Source: SISMIGRA for migration data.

Table III.27: Balance test on STI MAR dataset: Immigrants with Missing versus Not Missing Location

	(1)		(2)		T-test	Normalized
	Recorded MUN		No Recorded MUN		P-value	difference
Variable	N	Mean/SE	N	Mean/SE	(1)-(2)	(1)-(2)
F	5450	0.382	88502	0.205	0.000***	0.432
		(0.007)		(0.001)		
M	5450	0.618	88502	0.795	0.000***	-0.431
		(0.007)		(0.001)		
18-29	5450	0.443	88502	0.240	0.000***	0.468
		(0.007)		(0.001)		
30-39	5450	0.225	88502	0.480	0.000***	-0.510
		(0.006)		(0.002)		
40-49	5450	0.097	88502	0.193	0.000***	-0.245
		(0.004)		(0.001)		
50-59	5450	0.039	88502	0.046	0.010**	-0.037
		(0.003)		(0.001)		
60+	5450	0.013	88502	0.011	0.000***	0.018
		(0.002)		(0.000)		
Married	5450	0.172	2160	0.184	0.380	-0.031
		(0.005)		(0.008)		
Divorced	5450	0.010	2160	0.025	0.001***	-0.123
		(0.001)		(0.003)		
Separated	5450	0.001	2160	0.014	0.000***	-0.207
		(0.000)		(0.003)		
Single	5450	0.794	2160	0.743	0.000***	0.122
		(0.005)		(0.009)		
Civil Union	5450	0.020	2160	0.028	0.002**	-0.054
		(0.002)		(0.004)		
Widower	5450	0.003	2160	0.006	0.283	-0.032
		(0.001)		(0.002)		
Ative	5450	0.995	88502	0.484	0.000***	1.023
		(0.001)		(0.002)		
Excluded	5450	0.000	88502	0.001	0.013**	-0.017
		(0.000)		(0.000)		
Inative	5450	0.004	88502	0.515	0.000***	-1.022
		(0.001)		(0.002)		

Notes: This table presents the balance test between observable characteristics of immigrants with reported location of residence and those without. The table reports the T-test p-value and the normalised difference. STI-MAR for migration data.

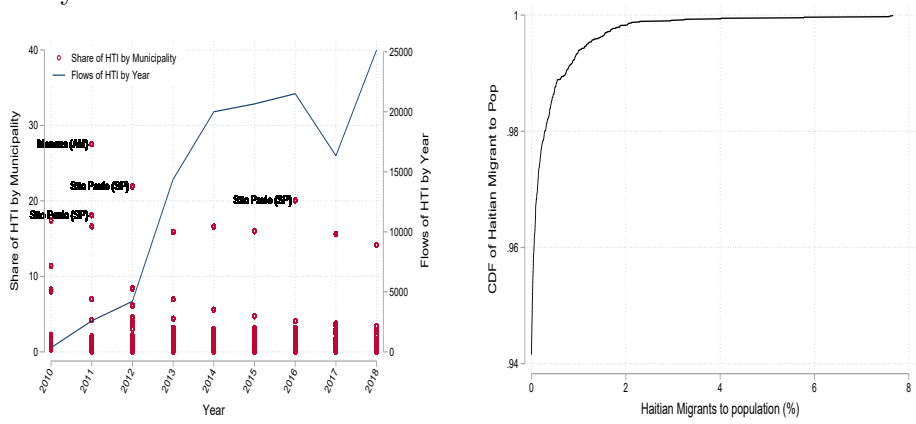
III.D Supplementary Material: Descriptive Analysis

The following section presents descriptive analysis of the context studied. Figure III.17 shows the shares of Haitian immigrants as a percent of total Haitian flows and the total flows of Haitians per year from 2010 to 2017. I also show the cumulative distribution function for the share of Haitian immigrants per population across municipalities in Brazil in 2018 (figure III.17(b)).

Figure III.17: Distribution of Haitian Immigrants

(a) The Distribution of Haitian Immigrants across Municipalities and Yearly Total Flows

(b) CDF of Haitian Migration to the Population



Notes: Figure III.17(a) depicts, for the Haitian immigration, the municipal share of each wave. The municipal share is calculated as $\frac{Mig_{m,t}}{Mig_t}$ where $Mig_{m,t}$ is the number of Haitian immigrants that located in municipality m at time t. The municipal share is then the ratio of the latter to total immigration flows of Haitians at time t. Figure III.17(b) shows the CDF of the main variable measuring the ratio of the Haitian concentration to the population in 2018, the year used for this analysis.

Further, in tables III.28 and III.29, I present additional descriptive statistics on the location of Haitian immigrants in Brazil per year.

Next, I present the variables used in the estimations as control variables (table III.31) and outcome variables (table III.30). The descriptive analysis is presented in the form of balance test analysis between Haitian rich municipalities (non-zero Haitian) and non-Haitian receiving municipalities (zero-Haitians).

Table III.28: Haitians' Municipality Characteristics per Year

	Percentage						Median	
	North (%)	North East (%)	Center-West (%)	South (%)	South East (%)	Border MUN (%)	Area	Pop Size
Year								
2010	17.6%	0.00%	5.9%	52.9%	23.5%	20.6%	6,564	1,802,014
2011	9.0%	0.00%	9.0%	52.2%	29.9%	11.9%	6,564	1,764,541
2012	7.6%	0.00%	10.2%	56.8%	25.4%	14.4%	1,521	446,911
2013	4.3%	0.39%	9.8%	57.6%	27.8%	15.7%	625	453,285
2014	4.1%	0.00%	10.0%	53.3%	32.6%	15.6%	539	339,979
2015	4.2%	0.00%	10.8%	53.1%	31.9%	16.0%	539	366,491
2016	3.8%	0.38%	9.4%	54.0%	32.5%	15.8%	651	477,798
2017	5.3%	0.00%	8.8%	53.7%	32.2%	17.2%	625	343,853
2018	4.4%	0.40%	9.3%	56.5%	29.4%	16.9%	625	324,476

Note: The table shows the number of municipalities that have Haitian immigrants. It also shows the distribution of these municipalities across regions (North, North-East, Center-West, South and South-East) as well as the percent as border municipalities. The table also shows the median area and population for these municipalities. Source: STI-MAR and SISMIGRA dataset.

Table III.29: Number of Flows and Visa Types of Haitian Immigrants

	Total	Mean Across Municipalities				
	Total Haitians Flows	Haitian Flows per Municipality	Asylum visa	Permanent visa	Temporary visa	Resident visa
Year						
2010	351	67.9	1.2	69.7	15.6	1.0
2011	2,586	376.6	29.8	369.5	7.6	1.2
2012	4,214	308.8	9.1	298.8	5.2	2.5
2013	14,388	599.0	3.4	567.2	13.8	22.9
2014	20,054	809.9	6.0	765.0	21.8	28.6
2015	20,676	771.6	4.8	700.5	47.6	28.9
2016	21,524	1,109.7	23.5	984.8	18.8	99.5
2017	16,339	590.5	97.9	6.9	81.0	440.6
2018	25,151	751.0	319.1		401.6	74.1

The table shows, by year, the flows of Haitian immigrants. It also shows in column 2, the average number of Haitians per municipality and the disaggregation across visa types. Source: STI-MAR and SISMIGRA dataset.

Table III.30: Electoral Outcomes: Balance tests between Haitian receiving and non-receiving municipalities

	(1)	(2)	T-test	Normalized
	Zero Haitian	Non-zero Haitian	P-value	difference
Variable	Mean/SE	Mean/SE	(1)-(2)	(1)-(2)
Turnout Ratio	72.789	78.284	0.286	-0.828
	(0.092)	(0.183)		
Presidential Left	56.715	42.980	0.057*	0.877
	(0.216)	(0.539)		
Presidential Center	43.164	56.871	0.055*	-0.876
	(0.215)	(0.540)		
Presidential Right	0.121	0.150	0.337	-0.255
	(0.002)	(0.008)		
Governor Left	37.048	35.788	0.009**	0.047
	(0.373)	(1.078)		
Governor Center	51.848	42.669	0.074*	0.276
	(0.463)	(1.612)		
Governor Right	11.104	21.543	0.222	-0.485
	(0.283)	(1.810)		
Senator Left	39.080	39.335	0.103	-0.014
	(0.255)	(0.778)		
Senator Center	36.754	39.775	0.150	-0.125
	(0.340)	(0.859)		
Senator Right	24.165	20.890	0.479	0.163
	(0.283)	(0.717)		
Federal Deputy Left	30.355	31.666	0.415	-0.073
	(0.249)	(0.901)		
Federal Deputy Center	31.774	32.117	0.350	-0.017
	(0.277)	(0.916)		
Federal Deputy Right	37.871	36.217	0.097*	0.089
	(0.259)	(0.798)		
State Deputy Left	31.483	35.205	0.009**	-0.203
	(0.253)	(0.961)		
State Deputy Center	34.614	34.351	0.056*	0.012
	(0.293)	(0.981)		
State Deputy Right	33.903	30.135	0.638	0.188
	(0.279)	(0.918)		

Note: The table shows the difference in means across different levels of electoral outcomes in 2010 (prior to the initial large immigration episodes). I test for statistically significant differences between municipalities that didn't receive any Haitian or Venezuelan immigrants, municipalities that received Haitian immigrants and municipalities that received Venezuelan immigrants. For the latter, the municipalities are mainly concentrated in the Northern bordering (to Venezuela) state of Roraima. Source: Tribunal Superior Eleitoral and SISMIGRA dataset.

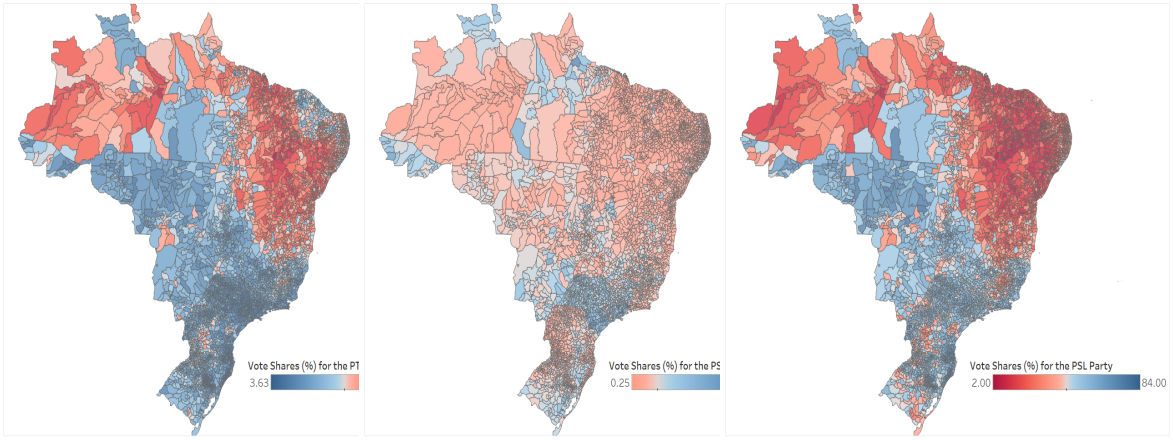
Table III.31: Control Variables: Balance tests between Haitian receiving and non-receiving municipalities

Variable	(1) Zero Haitian Mean/SE	(2) Non-zero Haitian Mean/SE	T-test P-value (1)-(2)	Normalized Difference (1)-(2)
Poverty Rate	0.243	0.056	0.000***	1.046
	(0.002)	(0.003)		
Sanitation Coverage	0.284	0.465	0.000***	-0.574
	(0.004)	(0.018)		
log(per Capita GDP)	8.543	9.348	0.000***	-1.157
	(0.009)	(0.029)		
log(Ratio of HH BF recipients)	1.220	0.114	0.000***	1.460
	(0.010)	(0.036)		
log(Value of BF transfer)	13.818	14.467	0.000***	-0.478
	(0.018)	(0.093)		
Public schools with water access	0.658	0.869	0.000***	-0.706
	(0.004)	(0.010)		
Water and Trash Coverage	0.522	0.730	0.000***	-0.964
	(0.003)	(0.009)		
People employed in Agriculture	0.376	0.118	0.000***	1.385
	(0.002)	(0.007)		
HH size	3.424	3.218	0.082*	0.466
	(0.006)	(0.014)		
Number of Rural Households	0.624	0.876	0.000***	-1.144
	(0.003)	(0.008)		
Homicides	0.000	0.000	0.000***	-0.199
	(0.000)	(0.000)		
Area of Municipality	1488.876	2098.886	0.000***	-0.109
	(71.919)	(559.465)		
Income ratio of White to Black	1.436	1.504	0.002**	-0.229
	(0.004)	(0.017)		
Income ratio of White to Brown	1.364	1.486	0.000***	-0.517
	(0.003)	(0.015)		
Income ratio of Brown to Black	0.967	0.995	0.363	-0.166
	(0.002)	(0.005)		
% of ppl under 5 years old	9.086	8.244	0.880	0.398
	(0.030)	(0.081)		
% of ppl between 5 and 14 years old	16.396	14.371	0.000***	0.677
	(0.041)	(0.111)		
% of ppl between 15 and 24 years old	17.884	17.607	0.000***	0.145
	(0.027)	(0.069)		
% of ppl between 25 and 39 years old	22.270	24.683	0.000***	-1.089
	(0.030)	(0.109)		
% of ppl between 40 and 59 years old	22.190	24.387	0.035**	-0.553
	(0.055)	(0.137)		
% of ppl more than 60 years old	12.174	10.707	0.000***	0.448
	(0.045)	(0.159)		
% of illiterate ppl less than 15 years old	16.828	5.362	0.000***	1.165
	(0.134)	(0.158)		
% of illiterate ppl between 25 and 39 years old	10.351	2.215	0.000***	0.942
	(0.120)	(0.091)		
% of illiterate ppl between 40 and 59 years old	21.859	6.142	0.000***	1.058
	(0.204)	(0.224)		
% of illiterate ppl more than 60 years old	42.138	18.839	0.000***	1.237
	(0.255)	(0.498)		
Income ratio male to femal	0.724	0.684	0.000***	0.479
	(0.001)	(0.003)		
HH with adequate sanitation	34.100	63.721	0.000***	-0.987
	(0.405)	(1.494)		
HH with inadequate sanitation	20.624	4.444	0.000***	0.992
	(0.224)	(0.428)		
HH up to 1/4 min wage	24.463	5.880	0.000***	1.071
	(0.238)	(0.330)		
HH up to 1/2 min wage	49.686	20.974	0.000***	1.308
	(0.295)	(0.611)		
HH up to 60 % of the median	42.476	15.879	0.000***	1.236
	(0.291)	(0.539)		
Ratio of Tax to GDP	0.060	0.127	0.000***	-1.556
	(0.001)	(0.003)		
Ratio of employed people to population	0.128	0.298	0.000***	-1.469
	(0.001)	(0.009)		
Ratio of paid people to population	0.104	0.249	0.000***	-1.413
	(0.001)	(0.009)		
Ratio of Companies to population	0.021	0.036	0.000***	-0.967
	(0.000)	(0.001)		
Value added agriculture divided by GDP	0.273	0.395	0.000***	-1.149
	(0.001)	(0.006)		
Value added Industry divided by GDP	0.126	0.258	0.000***	-0.965
	(0.002)	(0.006)		
Value added Services divided by GDP	0.332	0.143	0.000***	1.095
	(0.002)	(0.005)		
Gini Index	0.504	0.484	0.001**	0.301
	(0.001)	(0.004)		

Notes: This table shows the control variables used in the estimations of the results presented in this paper. In addition to those control variables, I include in the estimations, the characteristics of the electorate in 2014 as presented in table III.30. This table shows the overall mean of each variable across municipalities as well as the difference in means across these controls between Haitian rich municipalities in 2018 and those that didn't receive any Haitians.

Figure III.18: Votes for Haddad, Alckmin, and Bolsonaro in 2018

(a) Haddad (b) Alckmin (c) Bolsonaro



Note: Author's analysis of the elections data by municipality. The depicted voting outcome is equal to the ratio of votes going to each of the three Presidential candidates, Haddad, Alckmin and Bolsonaro, as a proportion to total votes in the presidential elections, per municipality. Source: Tribunal Superior Eleitoral.

Finally, in this section, I present in figure III.19 presents a visual depiction of the linguistic proximity to Haitian Creole for each country based on the main language spoken in the respective country.

Table III.32 – Language Assigned to Each Country

Country	Language Name	Language Code
Trinidad and Tobago, Republic of	Saint Lucian Creole French	acf
Mauritania, Islamic Republic of	Tunisian Arabic	aeb
United Arab Emirates	Gulf Arabic	afb
Namibia, Republic of	Afrikaans	afr
Jordan, Hashemite Kingdom of	South Levantine Arabic	ajp
Montenegro, Republic of	Gheg Albanian	aln
Ethiopia, Federal Democratic Republic of	Amharic	amh
Lebanon, Lebanese Republic	North Levantine Arabic	apc
Sudan, Republic of	Sudanese Arabic	apd
Algeria, People's Democratic Republic of	Algerian Arabic	arq
Morocco, Kingdom of	Moroccan Arabic	ary

Continued on next page

Table III.32 – *Continued from previous page*

Country	Language Name	Language Code
Egypt, Arab Republic of	Egyptian Arabic	arz
Libyan Arab Jamahiriya	Libyan Arabic	ayl
Yemen	Sanaani Arabic	ayn
Belarus, Republic of	Belarusian	bel
Zambia, Republic of	Bemba (Zambia)	bem
Bangladesh, People's Republic of	Bengali	ben
Vanuatu, Republic of	Bislama	bis
Bulgaria, Republic of	Bulgarian	bul
Andorra, Principality of	Catalan	cat
Czech Republic	Czech	ces
Guam	Chamorro	cha
Micronesia, Federated States of	Chuukese	chk
China, People's Republic of	Mandarin Chinese	cmn
Sao Tome and Principe, Democratic Republic of	Saotomense	cri
Denmark, Kingdom of	Danish	dan
Germany, Federal Republic of	German	deu
Bhutan, Kingdom of	Dzongkha	dzo
Estonia, Republic of	Estonian	ekk
Greece, Hellenic Republic	Modern Greek	ell
Belize	English	eng
Equatorial Guinea, Republic of	Ewondo	ewo
Gabon, Gabonese Republic	Fang (Equatorial Guinea)	fan
Fiji, Republic of the Fiji Islands	Fijian	fij
Finland, Republic of	Finnish	fin
France, French Republic	French	fra
Haiti, Republic of	Haitian	hat

Continued on next page

Table III.32 – *Continued from previous page*

Country	Language Name	Language Code
Nigeria, Federal Republic of	Hausa	hau
Israel, State of	Modern Hebrew	heb
India, Republic of	Hindi	hin
Hungary, Republic of	Hungarian	hun
Armenia, Republic of	Eastern Armenian	hye
Indonesia, Republic of	Indonesian	ind
Italy, Italian Republic	Italian	ita
Japan	Japanese	jpn
Mongolia	Halh Mongolian	khk
Cambodia, Kingdom of	Central Khmer	khm
Rwanda, Republic of	Kinyarwanda	kin
Korea, Democratic People's Republic of	Korean	kor
Sierra Leone, Republic of	Krio	kri
Lao People's Democratic Republic	Lao	lao
Lithuania, Republic of	Lithuanian	lit
Mauritius, Republic of	Morisyen	mfe
Macedonia, The Former Yugoslav Republic of	Macedonian	mkd
Malta, Republic of	Maltese	mlt
Burkina Faso	Mossi	mos
Myanmar, Union of	Burmese	mya
Netherlands, Kingdom of the	Dutch	nld
Norway, Kingdom of	Norwegian	nor
Nepal, State of	Nepali	npi
Iran, Islamic Republic of	Western Farsi	pes
Poland, Republic of	Polish	pol
Guinea-Bissau, Republic of	Portuguese	por

Continued on next page

Table III.32 – *Continued from previous page*

Country	Language Name	Language Code
Afghanistan, Islamic Republic of	Dari	prs
Bosnia and Herzegovina	Vlax Romani	rmy
Burundi, Republic of	Rundi	run
Russian Federation	Russian	rus
Chad, Republic of	Chadian Arabic	shu
Sri Lanka, Democratic Socialist Republic of	Sinhala	sin
Serbia, Republic of	Slovak	slk
Croatia, Republic of	Slovenian	slv
Zimbabwe, Republic of	Shona	sna
Somalia, Somali Republic	Somali	som
Lesotho, Kingdom of	Southern Sotho	sot
Chile, Republic of	Spanish	spa
Swaziland, Kingdom of	Swati	ssw
Saint Vincent and the Grenadines	Vincentian Creole English	svc
Sweden, Kingdom of	Swedish	swe
Kenya, Republic of	Swahili	swh
Timor-Leste, Democratic Republic of	Tetum	tet
Philippines, Republic of the	Tagalog	tgl
Thailand, Kingdom of	Thai	tha
Eritrea, State of	Tigrinya	tir
Bahamas, Commonwealth of the	Taino	tnq
Uzbekistan, Republic of	Turkmen	tuk
Turkey, Republic of	Turkish	tur
Ukraine	Ukrainian	ukr
Pakistan, Islamic Republic of	Urdu	urd
Vietnam, Socialist Republic of	Vietnamese	vie

Continued on next page

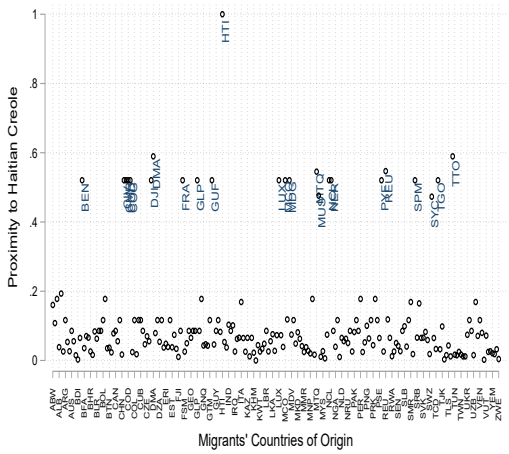
Table III.32 – *Continued from previous page*

Country	Language Name	Language Code
Comoros, Union of the	Ndzwani Comorian	wni
Senegal, Republic of	Wolof	wol
Malaysia	Colloquial Malay	zlm
South Africa, Republic of	Zulu	zul

Table III.32: Language Assigned by Country

Notes: This table shows the language assigned to each country. The table shows the language code as well as the full name of the language. Sources: CIA The World Factbook 2011, World Atlas Language Structures, and Automated Similarity Judgment Program.

Figure III.19: Proximity to Haitian Creole $\lambda_{f,creole}$ by Country



Notes: This figure shows the visual depiction of the linguistic proximity to Haitian Creole for each country. The linguistic measure is based on the Automated Similarity Judgement Program.

III.E Supplementary Material: Additional Results and Robustness Checks

This section presents additional results pertaining to the analysis, as well as robustness checks on the results. Each table has a note explaining the results shown.

Table III.33: Holding Turnout Ratio Fixed: Results across bands of turnout ratio

Across Similar Turnout Ratios

	(1) Geraldo Alckmin (PSDB)	(2) Fernando Haddad (PT)	(3) Jair Bolsonaro (PSL)	(4) Turnout Ratio (%)
HTI Flows (interval)/Pop 2010	0.586*	-0.747	0.844	-0.378
	(0.326)	(0.970)	(1.056)	(0.446)
	(1) Geraldo Alckmin (PSDB)	(2) Fernando Haddad (PT)	(3) Jair Bolsonaro (PSL)	(4) Turnout Ratio (%)
HTI Flows (interval)/Pop 2010	0.456***	0.00958	-1.053	-0.527
	(0.129)	(1.612)	(1.704)	(0.387)
Observations	1046	1046	1046	1046
Turnout Ratio Band	(85,95)	(85,95)	(85,95)	(85,95)

Notes: This table shows the outcome on vote share in the first round of elections for Alckmin, Haddad, and Bolsonaro, holding fixed the turnout ratio of municipalities within a certain band. As seen from the results, the Haitian immigration had no impact on turnout ratio, but had a positive increase of a magnitude similar to the one presented earlier for Alckmin vote shares. This is indicative evidence that the increase in vote shares to Alckmin could be due to vote switching. Data sources: IBGE 2010 Census data for control variables, Tribunal Superior Eleitoral and SISMIGRA for migration data.

Table III.34: Impact of Haitian Immigration on Vote Shares to Presidential Candidates (by Number of Voters)- 2SLS estimates

	(1) Fernando Haddad (PT)	(2) Jair Bolsonaro (PSL)	(3) Geraldo Alckmin (PSDB)	(4) Ciro Gomes (PDT)
HTI Flows (interval)/Pop 2010	-0.856	-0.302	0.341**	-0.390*
	(0.651)	(0.727)	(0.166)	(0.229)
Observations	5558	5558	5558	5558
Election Round	1	1	1	1
Mesoregion F.E	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes

	(1) Fernando Haddad (PT)	(2) Jair Bolsonaro (PSL)
HTI Flows (interval)/Pop 2010	-0.961	-0.0773
	(0.744)	(0.735)
Observations	5558	5558
Election Round	2	2
Mesoregion F.E	Yes	Yes
Controls	Yes	Yes

Notes: The table shows the 2SLS second stage estimates on presidential vote shares. In this specification, presidential vote shares are defined as votes cast to certain candidate as a ratio of the number of registered voters per municipality. Robust standard errors in parentheses clustered at the mesoregion level. Regressions include mesoregion fixed effects and municipality controls in 2010. Results are shown for the full sample of municipalities. Data sources: IBGE 2010 Census data for control variables, Tribunal Superior Eleitoral and STI-MAR and SISMIGRA for migration data.

Table III.37: Turnout Ratio: General

OLS Estimates						
	(1) Turnout Ratio (%)	(2) Temp. Transfer Turnout Ratio (%)	(3) Disable Turnout Ratio (%)	(4) Turnout Ratio (%)	(5) Temp. Transfer Turnout Ratio (%)	(6) Disable Turnout Ratio (%)
HTI Flows (interval)/Pop 2010	-0.163	0.00333	0.0392*	-0.148	0.00267	0.0356
	(0.173)	(0.00416)	(0.0227)	(0.210)	(0.00417)	(0.0221)
IV 2SLS Estimates						
	(1) Turnout Ratio (%)	(2) Temp. Transfer Turnout Ratio (%)	(3) Disable Turnout Ratio (%)	(4) Turnout Ratio (%)	(5) Temp. Transfer Turnout Ratio (%)	(6) Disable Turnout Ratio (%)
HTI Flows (interval)/Pop 2010	-0.970**	-0.00425	0.0343	-0.868*	-0.00372	0.0343
	(0.485)	(0.00634)	(0.0365)	(0.497)	(0.00654)	(0.0367)
Observations	5558	5558	5558	5558	5558	5558
Election Round	1	1	1	2	2	2
Mesoregion F.E	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The table show the OLS and 2SLS second stage estimates for the impact of migration on turnout ratios. The results are shown for all the sample of municipalities with available control data for the two rounds of elections in 2018. The outcome variable in columns (1) and (4) is $turnout ratio = \frac{votes_{casted}}{Voters_{registered}}$. The outcome variables in columns (2) and (5) are the temporary transfer turnout ratio and (3) and (6) are the turnout ratio for people with disability. The temporary transfer turnout ratio represents the ratio of voters that attended the election but asked that their vote is moved elsewhere, as a percent of all registered voters. This is usually done is a voter is registered

Table III.35: Impact of Haitian Immigration on Blank Votes - 2SLS estimates

Blank Votes to Total Voters (First Round)					
	(1)	(2)	(3)	(4)	(5)
	President	Governor	Senator	Federal Deputy	State Deputy
HTI Flows (interval)/Pop 2010	-0.0615	-0.0309	-0.233	0.0121	-0.0731
	(0.116)	(0.219)	(0.505)	(0.263)	(0.293)
Observations	5558	5558	5558	5558	5558
Mean in Haitian-rich Mun	2.029	3.593	7.879	2.947	2.813
Blank Votes to Casted Votes (First Round)					
	(1)	(2)	(3)	(4)	(5)
	President	Governor	Senator	Federal Deputy	State Deputy
HTI Flows (interval)/Pop 2010	-0.0434	0.0167	-0.109	0.0800	-0.0246
	(0.137)	(0.266)	(0.614)	(0.313)	(0.335)
Observations	5558	5558	5558	5558	5558
Mean in Haitian-rich Mun	2.508	4.449	9.707	3.631	3.464
Mesoregion F.E	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
Blank Votes to Total Voters (Second Round)					
	(1)	(2)			
	President	Governor			
HTI Flows (interval)/Pop 2010	0.00623	-0.136			
	(0.0624)	(0.104)			
Observations	5558	5558			
Mean in Haitian-rich Mun	2.029	3.593			
Blank Votes to Casted Votes (Second Round)					
	(1)	(2)			
	President	Governor			
HTI Flows (interval)/Pop 2010	0.0282	-0.124			
	(0.0733)	(0.139)			
Observations	5558	5558			
Mean in Haitian-rich Mun	2.508	4.449			
Mesoregion F.E	Yes	Yes			
Controls	Yes	Yes			

Notes: The table shows the 2SLS second stage estimates on blank votes cast across the different elections. Robust standard errors in parentheses clustered at the mesoregion level. Regressions include mesoregion fixed effects and municipality controls in 2010. Results are shown for the full sample of municipalities. Data sources: IBGE 2010 Census data for control variables, Tribunal Superior Eleitoral and STI-MAR and SISMIGRA for migration data.

Table III.36: Impact of Haitian Immigration on Null Votes - 2SLS estimates

Blank Votes to Total Voters (First Round)					
	(1)	(2)	(3)	(4)	(5)
	President	Governor	Senator	Federal Deputy	State Deputy
HTI Flows (interval)/Pop 2010	0.0532	0.171	0.276	-0.326	-0.351
	(0.152)	(0.356)	(0.722)	(0.364)	(0.462)
Observations	5558	5558	5558	5558	5558
Mean in Haitian-rich Mun	4.646	7.736	12.93	3.654	3.764
Blank Votes to Casted Votes (First Round)					
	(1)	(2)	(3)	(4)	(5)
	President	Governor	Senator	Federal Deputy	State Deputy
HTI Flows (interval)/Pop 2010	0.117	0.296	0.651	-0.334	-0.359
	(0.180)	(0.425)	(0.840)	(0.423)	(0.537)
Observations	5558	5558	5558	5558	5558
Mean in Haitian-rich Mun	5.760	9.600	15.96	4.511	4.626
Mesoregion F.E	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
Blank Votes to Total Voters (Second Round)					
	(1)	(2)			
	President	Governor			
HTI Flows (interval)/Pop 2010	0.170	-0.0968			
	(0.160)	(0.250)			
Observations	5558	5558			
Mean in Haitian-rich Mun	4.646	7.736			
Blank Votes to Casted Votes (Second Round)					
	(1)	(2)			
	President	Governor			
HTI Flows (interval)/Pop 2010	0.279	0.0122			
	(0.178)	(0.321)			
Observations	5558	5558			
Mean in Haitian-rich Mun	5.760	9.600			
Mesoregion F.E	Yes	Yes			
Controls	Yes	Yes			

Notes: The table shows the 2SLS second stage estimates on blank votes cast across the different elections. Robust standard errors in parentheses clustered at the mesoregion level. Regressions include mesoregion fixed effects and municipality controls in 2010. Results are shown for the full sample of municipalities. Data sources: IBGE 2010 Census data for control variables, Tribunal Superior Eleitoral and STI-MAR and SISMIGRA for migration data.

Table III.39: Turnout Ratio: By Education and Gender (First Round)

Compulsory Voting (First Round)						
	Total		**Female**		**Male**	
	(1)	(2)	(3)	(4)	(5)	(6)
	Up to Incomplete High School	At least Complete High School	Up to Incomplete High School	At least Complete High School	Up to Incomplete High School	At least Complete High School
HTI Flows (interval)/Pop 2010	-1.082**	-1.006*	-0.960*	-0.882*	-1.200**	-1.165*
	(0.527)	(0.587)	(0.524)	(0.506)	(0.540)	(0.703)
Voluntary Voting (First Round)						
	Total		**Female**		**Male**	
	(1)	(2)	(3)	(4)	(5)	(6)
	Up to Incomplete High School	At least Complete High School	Up to Incomplete High School	At least Complete High School	Up to Incomplete High School	At least Complete High School
HTI Flows (interval)/Pop 2010	-0.0968	0.0884	-0.212	-0.153	-0.118	0.348
	(1.525)	(1.073)	(1.762)	(1.215)	(1.242)	(1.162)
Observations	5558	5558	5558	5558	5558	5558
Mesoregion F.E	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The table show the second stage 2SLS estimates for the impact of migration on turnout ratios for voters for whom it is compulsory to vote (compulsory voters) and for others for whom it is voluntary to vote, disaggregated by gender and education levels. The first panel represents the results on turnout ratios for compulsory voters and the second panel represents results on turnout ratios for voluntary voters. The results are shown for all the sample of municipalities with available control data for the first round of elections in 2018. The outcome variable in columns (1) and (4) is $turnoutratio = \frac{votes_{cast}}{Voters_{registered}}$. The outcome variables in columns (2)-(6) represent the turnout ratio for females and males, disaggregated by education levels. The latter turnout ratio represents all the votes cast by females or males for each education classification as a proportion of all registered males or female voters, within that education category. Robust standard errors in parentheses clustered at the mesoregion level. Regressions include mesoregion fixed effects and municipality controls in 2010. Data sources: IBGE 2010 Census data for control variables, Tribunal Superior Eleitoral and STI-MAR and SISMIGRA for migration data.

III.E.1 Heterogeneous Impacts

Table III.38: Turnout Ratio: Compulsory versus Voluntary

	First Round Elections			Second Round Elections		
Compulsory: OLS						
	Total	Females	Males	Total	Females	Males
HTI Flows (interval)/Pop 2010	-0.252	-0.292*	-0.208	-0.217	-0.229	-0.203
	(0.179)	(0.159)	(0.204)	(0.233)	(0.227)	(0.243)
Compulsory: IV 2SLS						
	Total	Females	Males	Total	Females	Males
HTI Flows (interval)/Pop 2010	-1.026*	-0.920**	-1.132*	-0.897	-0.894*	-0.899
	(0.525)	(0.462)	(0.599)	(0.560)	(0.529)	(0.616)
Mean of Dep. Variable	83.56	85.13	81.93	81.16	82.40	79.89
Voluntary: OLS						
	Total	Females	Males	Total	Females	Males
HTI Flows (interval)/Pop 2010	0.329	0.245	0.397	0.309	0.218	0.399
	(0.449)	(0.511)	(0.421)	(0.389)	(0.420)	(0.356)
Voluntary: IV 2SLS						
	Total	Females	Males	Total	Females	Males
HTI Flows (interval)/Pop 2010	0.125	-0.143	0.299	0.154	0.0127	0.174
	(1.344)	(1.609)	(1.094)	(1.193)	(1.385)	(1.014)
Observations	5558	5558	5558	5558	5558	5558
Mean of Dep. Variable	57.09	54.36	59.94	57.12	53.87	60.47
Mesoregion F.E	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The table show the OLS and 2SLS second stage estimates for the impact of migration on turnout ratios for voters for whom it is compulsory to vote (compulsory voters) and for others for whom it is voluntary to vote. The first panel represents the results on turnout ratios for compulsory voters and the second panel represents results on turnout ratios for voluntary voters. The results are shown for all the sample of municipalities with available control data for the two rounds of elections in 2018. The outcome variable in columns (1) and (4) is $turnoutratio = \frac{votes_{cast}}{Voters_{registered}}$. The outcome variables in columns (2), (3), (5) and (6) represent the turnout ratio for females and males. The latter turnout ratio represents all the votes cast by females or males as a proportion of all registered males or female voters. Robust standard errors in parentheses clustered at the mesoregion level. Regressions include mesoregion fixed effects and municipality controls in 2010. Data sources: IBGE 2010 Census data for control variables, Tribunal Superior Eleitoral and STI-MAR and SISMIGRA for migration data.

Table III.40: Impact of Haitian Immigration on Turnout Ratio: By Education and Gender (Second Round)

Compulsory Voting (First Round)	**Total**		**Female**		**Male**	
	(1) Up to Incomplete High School	(2) At least Complete High School	(3) Up to Incomplete High School	(4) At least Complete High School	(5) Up to Incomplete High School	(6) At least Complete High School
HTI Flows (interval)/Pop 2010	-0.878	-0.935*	-0.824	-0.935*	-0.943	-0.952
	(0.561)	(0.566)	(0.563)	(0.533)	(0.585)	(0.646)
Observations	5558	5558	5558	5558	5558	5558
Mean in Haitian-rich Mun	82.11	87.66	83.41	88.09	80.88	87.09
Voluntary Voting (First Round)	**Total**		**Female**		**Male**	
	(1) Up to Incomplete High School	(2) At least Complete High School	(3) Up to Incomplete High School	(4) At least Complete High School	(5) Up to Incomplete High School	(6) At least Complete High School
HTI Flows (interval)/Pop 2010	-0.129	0.561	-0.0798	0.0549	-0.335	0.940
	(1.321)	(1.129)	(1.520)	(1.307)	(1.109)	(1.210)
Observations	5558	5558	5558	5558	5558	5558
Mean in Haitian-rich Mun	46.02	77.21	41.44	75.35	51.54	79.17
Mesoregion F.E	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The table show the second stage 2SLS estimates for the impact of migration on turnout ratios for voters for whom it is compulsory to vote (compulsory voters) and for others for whom it is voluntary to vote. The first panel represents the results on turnout ratios for compulsory voters and the second panel represents results on turnout ratios for voluntary voters. The results are shown for all the sample of municipalities with available control data for the two rounds of elections in 2018. The outcome variable in columns (1) and (4) is $turnoutratio = \frac{votes_{casted}}{Voters_{registered}}$. The outcome variables in columns (2), (3), (5) and (6) represent the turnout ratio for females and males. The latter turnout ratio represents all the votes cast by females or males as a proportion of all registered males or female voters. Robust standard errors in parentheses clustered at the mesoregion level. Regressions include mesoregion fixed effects and municipality controls in 2010. Data sources: IBGE 2010 Census data for control variables, Tribunal Superior Eleitoral and STI-MAR and SISMIGRA for migration data.

Table III.41: Heterogeneous Impacts on Presidential Votes

Urban Municipalities				
	(1)	(2)	(3)	(4)
	Fernando Haddad (PT)	Jair Bolsonaro (PSL)	Geraldo Alckmin (PSDB)	Ciro Gomes (PDT)
HTI Flows (interval)/Pop 2010	-0.118	-0.229	0.298	-0.339
	(0.731)	(0.818)	(0.215)	(0.312)
Municipalities	2395	2395	2395	2395
F-stat	27.11	27.11	27.11	27.11
Anderson-Rubin Test	0.877	0.790	0.152	0.268
Rural Municipalities				
	(1)	(2)	(3)	(4)
	Fernando Haddad (PT)	Jair Bolsonaro (PSL)	Geraldo Alckmin (PSDB)	Ciro Gomes (PDT)
HTI Flows (interval)/Pop 2010	-0.387	0.163	0.284	-0.441
	(0.711)	(0.732)	(0.237)	(0.329)
Municipalities	2327	2327	2327	2327
F-stat	23.61	23.61	23.61	23.61
Anderson-Rubin Test	0.601	0.830	0.219	0.151
Municipalities with Large Areas				
	(1)	(2)	(3)	(4)
	Fernando Haddad (PT)	Jair Bolsonaro (PSL)	Geraldo Alckmin (PSDB)	Ciro Gomes (PDT)
HTI Flows (interval)/Pop 2010	1.500	-3.091	1.462	0.430
	(3.693)	(3.889)	(1.250)	(1.320)
Municipalities	2727	2727	2727	2727
F-stat	18.10	18.10	18.10	18.10
Anderson-Rubin Test	0.688	0.421	0.226	0.756
Municipalities with Small Areas				
	(1)	(2)	(3)	(4)
	Fernando Haddad (PT)	Jair Bolsonaro (PSL)	Geraldo Alckmin (PSDB)	Ciro Gomes (PDT)
HTI Flows (interval)/Pop 2010	-0.818	0.344	0.649***	-0.540**
	(0.813)	(0.887)	(0.207)	(0.267)
Municipalities	2782	2782	2782	2782
F-stat	27.87	27.87	27.87	27.87
Anderson-Rubin Test	0.340	0.710	0.00382	0.0398
Municipalities with Low Inequality				
	(1)	(2)	(3)	(4)
	Fernando Haddad (PT)	Jair Bolsonaro (PSL)	Geraldo Alckmin (PSDB)	Ciro Gomes (PDT)
HTI Flows (interval)/Pop 2010	-1.018	0.591	0.390*	-0.280
	(0.708)	(0.676)	(0.204)	(0.220)
Municipalities	2785	2785	2785	2785
F-stat	44.56	44.56	44.56	44.56
Anderson-Rubin Test	0.172	0.407	0.0844	0.210
Municipalities with Left Party Win in 2010				
	(1)	(2)	(3)	(4)
	Fernando Haddad (PT)	Jair Bolsonaro (PSL)	Geraldo Alckmin (PSDB)	Ciro Gomes (PDT)
HTI Flows (interval)/Pop 2010	-2.021**	1.900*	0.658	-0.545
	(0.967)	(1.140)	(0.405)	(0.466)
Municipalities	3926	3926	3926	3926
F-stat	8.031	8.031	8.031	8.031
Anderson-Rubin Test	0.0145	0.0725	0.0180	0.218
Municipalities with BF Recipients > Median				
	(1)	(2)	(3)	(4)
	Fernando Haddad (PT)	Jair Bolsonaro (PSL)	Geraldo Alckmin (PSDB)	Ciro Gomes (PDT)
HTI Flows (interval)/Pop 2010	0.626	-1.818	1.343	0.473
	(3.221)	(3.137)	(1.407)	(1.359)
Municipalities	2784	2784	2784	2784
F-stat	3.522	3.522	3.522	3.522
Anderson-Rubin Test	0.861	0.643	0.193	0.765
Municipalities with High Migration Concentration				
	(1)	(2)	(3)	(4)
	Fernando Haddad (PT)	Jair Bolsonaro (PSL)	Geraldo Alckmin (PSDB)	Ciro Gomes (PDT)
HTI Flows (interval)/Pop 2010	0.114	-1.762**	0.560	0.142
	(0.693)	(0.897)	(0.367)	(0.331)
Municipalities	241	241	241	241
F-stat	35.32	35.32	35.32	35.32
Anderson-Rubin Test	0.900	0.107	0.194	0.741

Notes: The table shows the second stage 2SLS estimates for the impact of migration on Presidential candidates across different municipality characteristics. Each characteristic represents a subsample of municipalities that belong to that characteristic. Equation 5.1 is estimated for each of these subsamples. The Kleibergen-Paap F stat and Anderson-Rubin test are reported under each estimation. Results are shown for the first round of elections. Robust standard errors in parentheses clustered at the mesoregion level. Regressions include mesoregion fixed effects and municipality controls in 2010. Data sources: IBGE 2010 Census data for control variables, Tribunal Superior Eleitoral and STI-MAR and SISMIGRA for migration data.

Table III.42: Heterogeneous Impacts on Voter Turnout

Urban Municipalities			
	(1)	(2)	(3)
	Turnout Ratio	Female Turnout Ratio	Male Turnout Ratio
HTI Flows (interval)/Pop 2010	-0.472	-0.389	-0.572
	(0.406)	(0.370)	(0.484)
Municipalities	2395	2395	2395
F-stat	27.13	27.13	27.13
Anderson-Rubin Test	0.279	0.317	0.278
Rural Municipalities			
	(1)	(2)	(3)
	Turnout Ratio	Female Turnout Ratio	Male Turnout Ratio
HTI Flows (interval)/Pop 2010	-0.461	-0.419	-0.521
	(0.410)	(0.376)	(0.486)
Municipalities	2327	2327	2327
F-stat	23.62	23.62	23.62
Anderson-Rubin Test	0.302	0.296	0.330
Municipalities with Large Areas			
	(1)	(2)	(3)
	Turnout Ratio	Female Turnout Ratio	Male Turnout Ratio
HTI Flows (interval)/Pop 2010	-3.082	-3.010	-3.132
	(2.150)	(2.059)	(2.266)
Municipalities	2727	2727	2727
F-stat	18.10	18.10	18.10
Anderson-Rubin Test	0.109	0.109	0.118
Municipalities with Small Areas			
	(1)	(2)	(3)
	Turnout Ratio	Female Turnout Ratio	Male Turnout Ratio
HTI Flows (interval)/Pop 2010	-1.162**	-1.133**	-1.202*
	(0.542)	(0.501)	(0.616)
Municipalities	2782	2782	2782
F-stat	27.87	27.87	27.87
Anderson-Rubin Test	0.0361	0.0211	0.0641
Municipalities with Low Inequality			
	(1)	(2)	(3)
	Turnout Ratio	Female Turnout Ratio	Male Turnout Ratio
HTI Flows (interval)/Pop 2010	-0.776*	-0.767*	-0.788
	(0.466)	(0.413)	(0.538)
Municipalities	2785	2785	2785
F-stat	44.55	44.55	44.55
Anderson-Rubin Test	0.162	0.115	0.218
Mun with Left Party Win in 2010			
	(1)	(2)	(3)
	Turnout Ratio	Female Turnout Ratio	Male Turnout Ratio
HTI Flows (interval)/Pop 2010	-1.547*	-1.372*	-1.735*
	(0.844)	(0.767)	(0.935)
Municipalities	3926	3926	3926
F-stat	8.031	8.031	8.031
Anderson-Rubin Test	0.0451	0.0452	0.0459
Mun with BF Recipients > Median			
	(1)	(2)	(3)
	Turnout Ratio	Female Turnout Ratio	Male Turnout Ratio
HTI Flows (interval)/Pop 2010	-3.497*	-2.910*	-4.091*
	(1.958)	(1.680)	(2.306)
Municipalities	2784	2784	2784
F-stat	3.522	3.522	3.522
Anderson-Rubin Test	0.111	0.154	0.0859
Municipalities with High Migration Concentration			
	(1)	(2)	(3)
	Turnout Ratio	Female Turnout Ratio	Male Turnout Ratio
HTI Flows (interval)/Pop 2010	-0.00728	0.111	-0.138
	(0.638)	(0.595)	(0.719)
Municipalities	241	241	241
F-stat	35.33	35.33	35.33
Anderson-Rubin Test	0.993	0.887	0.884

Notes: The table shows the second stage 2SLS estimates for the impact of migration on turnout ratio for all voters, and for females and males, across different municipality characteristics. Each characteristic represents a subsample of municipalities that belong to that characteristic. Equation 5.1 is estimated for each of these subsamples. The Kleibergen-Paap F stat and Anderson-Rubin test are reported under each estimation. Results are shown for the first round of elections. Robust standard errors in parentheses clustered at the mesoregion level. Regressions include mesoregion fixed effects and municipality controls in 2010. Data sources: IBGE 2010 Census data for control variables, Tribunal Superior Eleitoral and STI-MAR and SISMIGRA for migration data.

Table III.43: Impact of Haitian Immigration on Heterogeneous Impacts on Presidential Votes based on Visa Status

Asylum						
	(1)	(2)	(3)	(4)	(5)	(6)
	Fernando Haddad (PT)	Jair Bolsonaro (PSL)	Geraldo Alckmin (PSDB)	Ciro Gomez (PDT)	Fernando Haddad (PT)	Jair Bolsonaro (PSL)
Asylum	-3.693	-25.76	4.530	-5.933*	-7.837	-23.71
	(8.602)	(32.45)	(2.928)	(3.127)	(9.159)	(35.43)
Municipalities Nonzero	157	157	157	157	157	157
F-stat	20.03	20.03	20.03	20.03	20.03	20.03
Anderson-Rubin Test	0.688	0.407	0.110	0.0495	0.428	0.489
Temporary and Resident Visa						
	(1)	(2)	(3)	(4)	(5)	(6)
	Fernando Haddad (PT)	Jair Bolsonaro (PSL)	Geraldo Alckmin (PSDB)	Ciro Gomez (PDT)	Fernando Haddad (PT)	Jair Bolsonaro (PSL)
Temporary and Resident	-1.240	-8.651	1.521*	-1.993	-2.632	-7.964
	(3.058)	(10.32)	(0.911)	(1.246)	(3.369)	(11.40)
Municipalities Nonzero	247	247	247	247	247	247
F-stat	16.59	16.59	16.59	16.59	16.59	16.59
Anderson-Rubin Test	0.688	0.407	0.110	0.0495	0.428	0.489
Permanent Visa						
	(1)	(2)	(3)	(4)	(5)	(6)
	Fernando Haddad (PT)	Jair Bolsonaro (PSL)	Geraldo Alckmin (PSDB)	Ciro Gomez (PDT)	Fernando Haddad (PT)	Jair Bolsonaro (PSL)
Permanent	-0.610	-4.258	0.749*	-0.981*	-1.295	-3.920
	(1.484)	(5.069)	(0.452)	(0.548)	(1.614)	(5.602)
Municipalities Nonzero	256	256	256	256	256	256
F-stat	28.04	28.04	28.04	28.04	28.04	28.04
Anderson-Rubin Test	0.688	0.407	0.110	0.0495	0.428	0.489

Notes: The table shows the second stage 2SLS estimates for the impact of migration on compulsory turnout ratio for all voters, and for females and males, across different municipality characteristics. Each characteristic represents a subsample of municipalities that belong to that characteristic. Equation 5.1 is estimated for each of these subsamples. The Kleibergen-Paap F stat and Anderson-Rubin test are reported under each estimation. Results are shown for the first round of elections. Robust standard errors in parentheses clustered at the mesoregion level. Regressions include mesoregion fixed effects and municipality controls in 2010. Data sources: IBGE 2010 Census data for control variables, Tribunal Superior Eleitoral and STI-MAR and SISMIGRA for migration data.

Table III.44: Heterogeneous Impacts of Haitian Immigration on Voter Turnout by Different Visa Holders

Asylum						
	(1)	(2)	(3)	(4)	(5)	(6)
	Turnout	Turnout Females	Turnout Males	Turnout	Turnout Females	Turnout Males
Asylum	-10.29	-9.224	-11.35	-8.993	-8.962	-9.014
	(6.672)	(5.886)	(7.537)	(6.932)	(6.545)	(7.517)
Municipalities Nonzero	157	157	157	157	157	157
F-stat	20.03	20.03	20.03	20.03	20.03	20.03
Anderson-Rubin Test	0.0629	0.0586	0.0706	0.127	0.108	0.162
Temporary and Resident Visa						
	(1)	(2)	(3)	(4)	(5)	(6)
	Turnout	Turnout Females	Turnout Males	Turnout	Turnout Females	Turnout Males
Temporary and Resident	-3.455**	-3.098**	-3.813**	-3.020*	-3.010*	-3.027
	(1.698)	(1.501)	(1.925)	(1.817)	(1.750)	(1.972)
Municipalities Nonzero	247	247	247	247	247	247
F-stat	16.59	16.59	16.59	16.59	16.59	16.59
Anderson-Rubin Test	0.0629	0.0586	0.0706	0.127	0.108	0.162
Permanent Visa						
	(1)	(2)	(3)	(4)	(5)	(6)
	Turnout	Turnout Females	Turnout Males	Turnout	Turnout Females	Turnout Males
Permanent	-1.701*	-1.525**	-1.877*	-1.486	-1.481*	-1.490
	(0.868)	(0.761)	(0.992)	(0.921)	(0.863)	(1.019)
Municipalities Nonzero	256	256	256	256	256	256
F-stat	28.04	28.04	28.04	28.04	28.04	28.04
Anderson-Rubin Test	0.0629	0.0586	0.0706	0.127	0.108	0.162

Notes: The table shows the 2SLS first stage estimates. Robust standard errors in parentheses clustered at the mesoregion level. Regressions include mesoregion fixed effects and municipality controls in 2010. Results are shown for two sample: full sample of municipalities and municipalities that received at least one migrant. The predicted Haitian migration variable is adjusted by population in 2010. Data sources: IBGE 2010 Census data for control variables, Tribunal Superior Eleitoral and SISMIGRA for migration data.

Table III.45: Impact of Haitian Immigration on Turnout Ratio: Heterogeneity by Education and Age

16-18 Years Old (First Round)								
	(1) Illiterate	(2) Literate	(3) Incomplete Primary	(4) Complete Primary	(5) Incomplete High School	(6) Complete High School	(7) Incomplete University	(8) University
HTI Flows (interval)/Pop 2010	-5.711	3.185	-1.210	-1.600	-0.639	0.763	-3.475	3.616
	(7.434)	(6.840)	(1.449)	(2.744)	(0.591)	(1.441)	(3.097)	(5.862)
Observations	5558	5558	5558	5558	5558	5558	5558	5558
Mean in Haitian-rich Mun	34.04	27.67	64.66	70.57	84.15	88.78	90.41	37.91
19-39 Years Old (First Round)								
	(1) Illiterate	(2) Literate	(3) Incomplete Primary	(4) Complete Primary	(5) Incomplete High School	(6) Complete High School	(7) Incomplete University	(8) University
HTI Flows (interval)/Pop 2010	-0.248	1.398	-1.897**	-1.821*	-1.401*	-1.122	-0.770	-0.813*
	(1.961)	(2.448)	(0.749)	(0.964)	(0.788)	(0.844)	(0.839)	(0.430)
Observations	5558	5558	5558	5558	5558	5558	5558	5558
Mean in Haitian-rich Mun	53.85	73.30	72.26	75.36	79.60	85.09	85.62	88.73
40-59 Years Old (First Round)								
	(1) Illiterate	(2) Literate	(3) Incomplete Primary	(4) Complete Primary	(5) Incomplete High School	(6) Complete High School	(7) Incomplete University	(8) University
HTI Flows (interval)/Pop 2010	1.085	-1.831*	-0.680**	-0.640	-0.814	-0.949**	-0.602	-0.905*
	(1.686)	(1.027)	(0.342)	(0.472)	(0.646)	(0.402)	(0.665)	(0.465)
Observations	5558	5558	5558	5558	5558	5558	5558	5558
Mean in Haitian-rich Mun	63.19	84.47	86.17	87.70	87.56	90.15	88.73	91.73
70+ Years Old (First Round)								
	(1) Illiterate	(2) Literate	(3) Incomplete Primary	(4) Complete Primary	(5) Incomplete High School	(6) Complete High School	(7) Incomplete University	(8) University
HTI Flows (interval)/Pop 2010	0.344	-1.826	0.527	-0.237	1.714	-0.499	-1.119	2.796
	(1.492)	(2.197)	(1.867)	(1.925)	(3.774)	(1.715)	(4.637)	(1.869)
Observations	5558	5558	5558	5558	5558	5558	5558	5558
Mean in Haitian-rich Mun	16.33	32.05	43.75	53.27	59.68	65.52	66.99	74.04
Mesoregion F.E	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The table shows the second stage 2SLS estimates for the impact of migration on turnout ratios for voters across four different age groups: 16-18 and 70 + (where it is voluntary for them to vote) and 19-39 and 40-59 where it is compulsory to vote. The age categories are further subdivided into 8 education categories as specified by the Tribunal Superior Eleitoral. It is voluntary to vote for illiterate voters. Results are shown for the first round of elections. Robust standard errors in parentheses clustered at the mesoregion level. Regressions include mesoregion fixed effects and municipality controls in 2010. Data sources: IBGE 2010 Census data for control variables, Tribunal Superior Eleitoral and STI-MAR and SISMIGRA for migration data.

Table III.46: Impact of Haitian Immigration on Turnout Ratio for Females: By Education and Age for Females

Females (16-18 Age)								
	(1) Illiterate	(2) Literate	(3) Incomplete Primary	(4) Complete Primary	(5) Incomplete High School	(6) Complete High School	(7) Incomplete University	(8) University
HTI Flows (interval)/Pop 2010	-1.522	0.982	0.655	-6.219	-1.198*	-0.288	-2.398	5.899
	(6.686)	(2.860)	(2.024)	(6.938)	(0.695)	(2.380)	(3.639)	(6.359)
Observations	5558	5558	5558	5558	5558	5558	5558	5558
Mean in Haitian-rich Mun	22.47	16.96	63.91	66.08	83.73	88.36	89.06	29.07
Females (19-39 Age)								
	(1) Illiterate	(2) Literate	(3) Incomplete Primary	(4) Complete Primary	(5) Incomplete High School	(6) Complete High School	(7) Incomplete University	(8) University
HTI Flows (interval)/Pop 2010	-9.533*	-0.406	-1.244	-2.604***	-1.191	-0.796	-0.887	-0.625
	(5.121)	(3.828)	(1.092)	(0.660)	(0.845)	(0.795)	(0.657)	(0.475)
Observations	5558	5558	5558	5558	5558	5558	5558	5558
Mean in Haitian-rich Mun	52.97	74.21	74.07	76.63	80.21	85.36	85.90	89.22
Females (40-69 Age)								
	(1) Illiterate	(2) Literate	(3) Incomplete Primary	(4) Complete Primary	(5) Incomplete High School	(6) Complete High School	(7) Incomplete University	(8) University
HTI Flows (interval)/Pop 2010	-0.212	-1.171	-0.788**	-0.125	-0.752	-1.051***	-0.692	-0.837*
	(1.584)	(0.828)	(0.359)	(0.424)	(0.514)	(0.396)	(0.774)	(0.463)
Observations	5558	5558	5558	5558	5558	5558	5558	5558
Mean in Haitian-rich Mun	64.62	85.58	87.22	88.68	88.12	90.56	89.05	92.37
Females (70+)								
	(1) Illiterate	(2) Literate	(3) Incomplete Primary	(4) Complete Primary	(5) Incomplete High School	(6) Complete High School	(7) Incomplete University	(8) University
HTI Flows (interval)/Pop 2010	-0.278	-1.079	0.0946	-0.694	-5.709	-0.172	1.002	3.663
	(1.299)	(2.447)	(1.978)	(2.539)	(7.621)	(1.850)	(7.983)	(3.079)
Observations	5558	5558	5558	5558	5558	5558	5558	5558
Mean in Haitian-rich Mun	14.55	28.61	39.02	48.41	53.72	61.09	58.78	71.56
Mesoregion F.E	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The table shows the second stage 2SLS estimates for the impact of migration on turnout ratios for female voters across four different age groups: 16-18 and 70 + (where it is voluntary for them to vote) and 19-39 and 40-59 where it is compulsory to vote. The age categories are further subdivided into 8 education categories as specified by the Tribunal Superior Eleitoral. It is voluntary to vote for illiterate voters. Results are shown for the first round of elections. Robust standard errors in parentheses clustered at the mesoregion level. Regressions include mesoregion fixed effects and municipality controls in 2010. Data sources: IBGE 2010 Census data for control variables, Tribunal Superior Eleitoral and STI-MAR and SISMIGRA for migration data.

Table III.47: Impact of Haitian Immigration on Turnout Ratio for Males: By Education and Age for Males

Males (16-18 Age) By Education Level (First Round)								
	(1) Illiterate	(2) Literate	(3) Incomplete Primary	(4) Complete Primary	(5) Incomplete High School	(6) Complete High School	(7) Incomplete University	(8) University
HTI Flows (interval)/Pop 2010	-5.721	5.923	-1.563	-1.966	-0.133	0.463	-3.878	0.212
	(4.354)	(7.883)	(2.288)	(2.972)	(0.621)	(1.730)	(4.127)	(4.215)
Observations	5558	5558	5558	5558	5558	5558	5558	5558
Mean in Haitian-rich Mun	28.18	20.19	64.47	70.27	84.60	88.34	87.76	25.05
Males (19-39 Age) By Education Level (First Round)								
	(1) Illiterate	(2) Literate	(3) Incomplete Primary	(4) Complete Primary	(5) Incomplete High School	(6) Complete High School	(7) Incomplete University	(8) University
HTI Flows (interval)/Pop 2010	3.715	2.568	-2.259***	-1.411	-1.592**	-1.466	-0.734	-1.085*
	(3.059)	(3.304)	(0.657)	(1.334)	(0.789)	(0.958)	(1.144)	(0.584)
Observations	5558	5558	5558	5558	5558	5558	5558	5558
Mean in Haitian-rich Mun	53.38	70.84	70.82	74.26	79.05	84.74	85.20	87.83
Males (40-59 Age) By Education Level (First Round)								
	(1) Illiterate	(2) Literate	(3) Incomplete Primary	(4) Complete Primary	(5) Incomplete High School	(6) Complete High School	(7) Incomplete University	(8) University
HTI Flows (interval)/Pop 2010	3.432**	-2.710**	-0.594*	-1.133*	-0.828	-0.794*	-0.666	-1.141*
	(1.721)	(1.373)	(0.354)	(0.615)	(0.864)	(0.464)	(0.843)	(0.588)
Observations	5558	5558	5558	5558	5558	5558	5558	5558
Mean in Haitian-rich Mun	61.39	83.24	85.09	86.77	87.01	89.66	88.39	90.69
Males (70+) By Education Level (First Round)								
	(1) Illiterate	(2) Literate	(3) Incomplete Primary	(4) Complete Primary	(5) Incomplete High School	(6) Complete High School	(7) Incomplete University	(8) University
HTI Flows (interval)/Pop 2010	1.716	-1.506	0.783	1.020	8.178***	0.773	1.549	4.413**
	(4.809)	(2.918)	(1.696)	(1.989)	(2.801)	(2.131)	(4.885)	(1.833)
Observations	5558	5558	5558	5558	5558	5558	5558	5558
Mean in Haitian-rich Mun	19.25	36.30	49.22	58.45	63.64	70.69	63.89	76.04
Mesoregion F.E	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The table shows the second stage 2SLS estimates for the impact of migration on turnout ratios for male voters across four different age groups: 16-18 and 70 + (where it is voluntary for them to vote) and 19-39 and 40-59 where it is compulsory to vote. The age categories are further subdivided into 8 education categories as specified by the Tribunal Superior Eleitoral. It is voluntary to vote for illiterate voters. Results are shown for the first round of elections. Robust standard errors in parentheses clustered at the mesoregion level. Regressions include mesoregion fixed effects and municipality controls in 2010. Data sources: IBGE 2010 Census data for control variables, Tribunal Superior Eleitoral and STI-MAR and SISMIGRA for migration data.

Table III.48: Turnout Ratio: Heterogeneity by Sex and Age

Female by Age-Range (First Round)				
	(1)	(2)	(3)	(4)
	Age 16-18	Age 19-39	Age 40-69	Age 70+
HTI Flows (interval)/Pop 2010	-0.978	-1.078	-0.786**	-0.214
	(0.776)	(0.661)	(0.336)	(1.767)
Observations	5558	5558	5558	5558
Female by Age-Range (Second Round)				
	(1)	(2)	(3)	(4)
	Age 16-18	Age 19-39	Age 40-69	Age 70+
HTI Flows (interval)/Pop 2010	-0.859	-1.064	-0.749*	-0.0764
	(1.120)	(0.721)	(0.410)	(1.519)
Observations	5558	5558	5558	5558
Mean in Haitian-rich Mun	84.12	82.68	88.18	36.86
Male by Age-Range (First Round)				
	(1)	(2)	(3)	(4)
	Age 16-18	Age 19-39	Age 40-69	Age 70+
HTI Flows (interval)/Pop 2010	-0.442	-1.496**	-0.659	-0.138
	(0.549)	(0.757)	(0.453)	(1.273)
Observations	5558	5558	5558	5558
Mean in Haitian-rich Mun	83.54	80.23	86.42	47.73
Male by Age-Range (Second Round)				
	(1)	(2)	(3)	(4)
	Age 16-18	Age 19-39	Age 40-69	Age 70+
HTI Flows (interval)/Pop 2010	-0.876	-1.078	-0.640	-0.145
	(0.918)	(0.834)	(0.428)	(1.192)
Observations	5558	5558	5558	5558
Mean in Haitian-rich Mun	83.54	80.23	86.42	47.73
Mesoregion F.E	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes

Notes: The table shows the 2SLS first stage estimates. Robust standard errors in parentheses clustered at the mesoregion level. Regressions include mesoregion fixed effects and municipality controls in 2010. Results are shown for two sample: full sample of municipalities and municipalities that received at least one migrant. The predicted Venezuelan migration variable is adjusted by population in 2010. Data sources: IBGE 2010 Census data for control variables, Tribunal Superior Eleitoral and SISMIGRA for migration data.

III.E.2 Varying the Sample Analysis

The following estimations present the 2SLS estimations of the impact of Haitian immigration on vote shares for Presidential candidates, but varying the municipality samples used for the analysis. I test two samples: the first excludes all municipalities that received other immigrants (between 2010-2018) but not Haitians; while the second, excludes all municipalities that received immigrants, and keep only those that received Haitians and those that didn't receive any migrant (between 2010-2018).

Table III.49: Sample of municipalities that received Haitian immigrants and those that received zero immigrants

First Round							
	(1)	(2)	(3)	(4)	(5)	(6)	
	Jair Bolsonaro (PSL)	Fernando Haddad (PT)	Marina Silva (REDE)	Geraldo Alckmin (PSDB)	Ciro Gomes (PDT)	Jose Maria Eymael (DC)	
HTI Flows (interval)/Pop 2010	-0.0765	-0.523	-0.0289	0.429*	-0.170	0.000520	
	(0.874)	(0.807)	(0.0258)	(0.221)	(0.271)	(0.00413)	
Observations	5139	5139	5139	5139	5139	5139	
Kleibergen-Paap F-stat	54.1444	54.1444	54.1444	54.1444	54.1444	54.1444	
Anderson-Rubin Test	0.9318	0.5295	0.3248	0.0853	0.5394	0.9011	
Cragg-Donald F Stat	962.6796	962.6796	962.6796	962.6796	962.6796	962.6796	
Election Round	1	1	1	1	1	1	
Mesoregion F.E	Yes	Yes	Yes	Yes	Yes	Yes	
Controls	Yes	Yes	Yes	Yes	Yes	Yes	
First Round							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Henrique Meirelles (MDB)	Joao Amoedo (NOVO)	Cabo Daciolo (PATRI)	Alvaro Dias (PODE)	Joao Vicente Goulart (PPL)	Guilherme Boulos (PSOL)	Vera Lucia (PSTU)
HTI Flows (interval)/Pop 2010	0.0921	0.240*	-0.0502	0.0802	-0.000937	0.00765	-0.0000486
	(0.0738)	(0.142)	(0.0415)	(0.0675)	(0.00312)	(0.0161)	(0.00346)
Observations	5139	5139	5139	5139	5139	5139	5139
Kleibergen-Paap F-stat	54.1444	54.1444	54.1444	54.1444	54.1444	54.1444	54.1444
Anderson-Rubin Test	0.1936	0.0793	0.2789	0.2354	0.7638	0.6319	0.9891
Cragg-Donald F Stat	962.6796	962.6796	962.6796	962.6796	962.6796	962.6796	962.6796
Election Round	1	1	1	1	1	1	1
Mesoregion F.E	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Second Round							
	(1)	(2)					
	Jair Bolsonaro (PSL)	Fernando Haddad (PT)					
HTI Flows (interval)/Pop 2010	0.389	-0.389					
	(0.937)	(0.937)					
Observations	5139	5139					
Kleibergen-Paap F-stat	54.1444	54.1444					
Anderson-Rubin Test	0.6881	0.6881					
Cragg-Donald F Stat	962.6796	962.6796					
Election Round	2	2					
Mesoregion F.E	Yes	Yes					
Controls	Yes	Yes					

Notes: The table shows the 2SLS second stage estimates for ideological subcategories (far-left, leftwing, centre, rightwing and far-right). Robust standard errors in parentheses clustered at the mesoregion level. Regressions include mesoregion fixed effects and municipality controls in 2010. Results are shown for the full sample of municipalities in the first round of the elections. Data sources: IBGE 2010 Census data for control variables, Tribunal Superior Eleitoral and STI-MAR and SISMIGRA for migration data.

Table III.50: Sample with only Haitian receiving municipalities and non-migrant receiving municipalities

First Round							
	(1)	(2)	(3)	(4)	(5)	(6)	
	Jair Bolsonaro (PSL)	Fernando Haddad (PT)	Marina Silva (REDE)	Geraldo Alckmin (PSDB)	Ciro Gomes (PDT)	Jose Maria Eymael (DC)	
HTI Flows (interval)/Pop 2010	1.461***	-1.879**	-0.0465*	0.736***	-0.894***	0.00132	
	(0.482)	(0.765)	(0.0280)	(0.219)	(0.192)	(0.00396)	
Observations	4927	4927	4927	4927	4927	4927	
Kleibergen-Paap F-stat	58.7271	58.7271	58.7271	58.7271	58.7271	58.7271	
Anderson-Rubin Test	0.0005	0.0077	0.1565	0.0095	0.0000	0.7417	
Cragg-Donald F Stat	550.7983	550.7983	550.7983	550.7983	550.7983	550.7983	
Election Round	1	1	1	1	1	1	
Mesoregion F.E	Yes	Yes	Yes	Yes	Yes	Yes	
Controls	Yes	Yes	Yes	Yes	Yes	Yes	
First Round							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Henrique Meirelles (MDB)	Joao Amoedo (NOVO)	Cabo Daciolo (PATRI)	Alvaro Dias (PODE)	Joao Vicente Goulart (PPL)	Guilherme Boulos (PSOL)	Vera Lucia (PSTU)
HTI Flows (interval)/Pop 2010	0.153	0.449**	-0.103*	0.109	-0.000860	0.0146*	0.000404
	(0.100)	(0.222)	(0.0553)	(0.0922)	(0.00735)	(0.00779)	(0.00456)
Observations	4927	4927	4927	4927	4927	4927	4927
Kleibergen-Paap F-stat	58.7271	58.7271	58.7271	58.7271	58.7271	58.7271	58.7271
Anderson-Rubin Test	0.0811	0.0403	0.1353	0.2166	0.9089	0.0761	0.9317
Cragg-Donald F Stat	550.7983	550.7983	550.7983	550.7983	550.7983	550.7983	550.7983
Election Round	1	1	1	1	1	1	1
Mesoregion F.E	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Second Round							
	(1)	(2)					
	Jair Bolsonaro (PSL)	Fernando Haddad (PT)					
HTI Flows (interval)/Pop 2010	2.017***	-2.017***					
	(0.598)	(0.598)					
Observations	4927	4927					
Kleibergen-Paap F-stat	58.7271	58.7271					
Anderson-Rubin Test	0.0003	0.0003					
Cragg-Donald F Stat	550.7983	550.7983					
Election Round	2	2					
Mesoregion F.E	Yes	Yes					
Controls	Yes	Yes					

Notes: The table shows the 2SLS second stage estimates for ideological subcategories (far-left, leftwing, centre, rightwing and far-right). Robust standard errors in parentheses clustered at the mesoregion level. Regressions include mesoregion fixed effects and municipality controls in 2010. Results are shown for the full sample of municipalities in the first round of the elections. Data sources: IBGE 2010 Census data for control variables, Tribunal Superior Eleitoral and STI-MAR and SISMIGRA for migration data.

III.E.3 Channels of Impact

III.E.3.1 Alternative measures on Natives' Flight

The second measure represents the yearly *change* in internal migration of natives as a proportion to the population, calculated as such:

$$\Delta \frac{InternalMigration}{Population_{t-1}}_{m,t} = \frac{InternalMigration}{Population}_{m,t} - \frac{InternalMigration}{Population}_{m,t-1} \tag{III.E.18}$$

The third measure represents the change in internal migration of natives relative to the base year 2010, also as a proportion of population, calculated as such:

$$\frac{InternalMigration}{Population_{2010}}_{m} = (Pop_{m,t} - Pop_{m,2011}) - \Big[(births_{m,t} - deaths_{m,t}) - (births_{m,2011} - deaths_{m,2011})\Big] \tag{III.E.19}$$

and finally the fourth measure represents the net native immigration flows summed across all years from 2010 to 2018, and calculated as such:

$$\frac{\sum_{t=2010}^{2018}(InternalMigration)}{Population_{2010}} \tag{III.E.20}$$

I present the results across all these measures in table III.51. I find that the Haitian immigration flow had no impact on any of the measures of natives' internal movement. This result is not surprising, after all, Haitian immigrant ratio to the municipal population had a mean[76] of 0.3% with a maximum of 7.2%. This small proportion is not enough for a native to move their residence elsewhere.

[76]This mean is calculated among municipalities that received at least one Haitian immigrant.

Table III.51: Channels of Impact: Natives' Internal Migration

Native's Internal Migration

	(1) $\frac{InternalMigration}{Pop_{t-1}}$	(2) $\Delta(\frac{InternalMigration}{Pop_{t-1}})$	(3) $\Delta(\frac{InternalMigration}{Population})_{2018-2010}$	(4) $\frac{\sum_{t=2010}^{2018}(InternalMigration)}{Population_{2010}}$
HTI Flows (interval)/Pop 2010	0.00116	-0.000284	0.0196	0.0177
	(0.00283)	(0.00223)	(0.0181)	(0.0148)
Observations	5242	5233	5135	5558
Mean of Outcome Variable	-0.00473	-0.00425	0.0638	0.0572
Mesoregion F.E	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes

Notes: The table shows the 2SLS second stage estimates of the impact of Haitian immigration on internal native immigration. Robust standard errors in parentheses clustered at the mesoregion level. Regressions include mesoregion fixed effects and municipality controls in 2010. Results are shown for the full sample of municipalities for which data on control variables is available. The predicted Haitian migration variable is adjusted by population in 2010. The mean of the dependent variable is shown for municipalities that had a Haitian immigrant in 2018. The dependent variables in this table are four different variations of the measurement on native's internal immigration. The dependent variable in column 1 shows yearly internal migration of natives as a proportion to the population per municipality in 2018, while that in column 2 shows the yearly *change* in internal migration of natives as a proportion to the population, and that in the third column shows the change in internal migration of natives relative to the base year 2010, also as a proportion of population. The fourth column represents the summation of all net native immigration flows from 2010 to 2018. Source: IBGE 2010 Census data for control variables, Tribunal Superior Eleitoral and SISMIGRA for migration data.

Table III.52: Descriptive Statistics on Presidential Turnout by Individual's Preferences on Immigration Policies

	Did you vote in the last presidential elections in 2018?		
	No	**Yes**	**Total**
	%	%	%
Sex			
Female	23.3	76.7	100.0
Male	24.9	75.1	100.0
Total	24.1	75.9	100.0
age			
16-	71.8	28.2	100.0
19-	21.3	78.7	100.0
40-	13.8	86.2	100.0
70-	48.9	51.1	100.0
Total	24.1	75.9	100.0
Area			
Capital (metro area)	22.2	77.8	100.0
Large-size city	25.2	74.8	100.0
Mid-size city	25.1	74.9	100.0
Small town	21.0	79.0	100.0
Rural area	24.1	75.9	100.0
Total	24.1	75.9	100.0
Income Range			
High (>500k)	25.1	74.9	100.0
Middle (between 50k and 500k)	25.0	75.0	100.0
Low (< 50k)	22.0	78.0	100.0
Total	24.1	75.9	100.0

Notes: The table shows descriptive statistics based on the Presidential political participation of Brazilians in 2018, voter individual characteristics as well as their attitudes towards immigrants. The survey asks directly about Venezuelans, Spanish and all other immigrants (not specifically about Haitians). The numbers in the table represent the percent of people, weighted by the sampling weights, that belonged to each category. Source: LAPOP dataset 2019.

III.E.3.2 Impact on Candidate Characteristics

Table III.53: Impact on Candidate Characteristics

	(1) Total Candidates	(2) White	(3) Black	(4) Brown	(5) Incomplete Primary	(6) Complete Primary	(7) Incomplete High School	(8) Complete High School	(9) Incomplete University	(10) University	(11) Male	(12) Female
IV First Stage: Haitian Migration												
HTI Flows (interval)/Pop 2010	13.44	10.97	1.489	0.801	0.164	0.623	0.369	2.561	1.330	8.351	8.433	5.012
	(17.07)	(12.53)	(2.011)	(2.470)	(0.323)	(0.699)	(0.437)	(3.914)	(1.663)	(9.972)	(10.95)	(6.114)
Observations	5558	5558	5558	5558	5558	5558	5558	5558	5558	5558	5558	5558
No of Mun with Candidates	1613	1287	473	949	303	420	250	1005	479	1179	1442	980
Municipality Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mesoregion F.E	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The table shows the 2SLS estimates for the impact of immigration on candidate characteristics (Race, Education and Gender). The results are shown for all the sample of municipalities with available control data. Outcome variable is the number of candidates with a certain characteristic that ran for election in a certain municipality. The latter represents the municipality of birth of the candidate. Robust standard errors in parentheses clustered at the mesoregion level. Regressions include mesoregion fixed effects and municipality controls in 2010. Data sources: IBGE 2010 Census data for control variables, Tribunal Superior Eleitoral and STI-MAR and SISMIGRA for migration data.

III.F Supplementary Material: Case Study of Venezuelan Immigration

The Venezuelan immigration is found to be concentrated in mainly two municipalities of entry, with potential spillover to other nearby areas within the state of Roraima (based on my analysis of SISMIGRA and STI-MAR migration datasets). The main challenge in estimating a causal relationship between the Venezuelan immigration and political outcomes is the absence of an adequate counterfactual that can be compared to the immigrant receiving municipalities. The presence of this counterfactual helps to answer the question of how would the voting outcomes be in the impacted areas had the Venezuelan immigration didn't occur. The affected municipalities due to the Venezuelan immigration aren't only small in number (two main ones, Pacaraima and Boa Vista, received more than 80% of all Venezuelan immigrants), but they could also be inherently different from non-affected municipalities. Therefore an exercise of calculating a difference in means will not be sufficient nor convincing as a causal estimate.

In order to identify the impact of the Venezuelan immigration on political outcomes, I employ a synthetic control estimation strategy (Abadie and Gardeazabal, 2003 ; Abadie et al., 2010) to create for each impacted area, a counterfactual municipality. The usage of this technique is motivated by having two municipalities Pacaraima and Boa Vista, the two most impacted municipalities. I depict the share of Venezuelan immigrants by municipality across the years in figure III.20(a). These two municipalities lie on a main road connecting Venezuela to Brazil. Pacaraima is the first municipality that an immigrant taking this road will reach and Boa Vista is the final destination within the state of Roraima (See map III.20(b)). The state of Roraima, of which Venezuelan immigrants constitute a fifth of the population, created the Federal Committee of Emergency Assistance, as well as refugee camps to assist the Venezuelans crossing its borders and who are in vulnerable conditions. In order to do so, I use pre-treatment (pre-2016) voting outcomes to create synthetic municipalities that would be an identical counterfactual to the affected municipalities (14 in total in the State of Roraima), and would serve as an adequate comparison for when after the Venezuelan immigration takes place. This technique is suitable in my context because the number of impacted municipalities is small and wouldn't be sufficient for a consistent estimation of other techniques. With the shortcomings of the data mentioned before, the technique will also allow me to ameliorate the complexity around migrant

reported locations: I do not only employ this technique on the two municipalities of reported entry into Brazil but also on other municipalities in the state of Roraima. I have a small number of years in my sample. The data at the municipal level has 6 years of electoral results available, 5 of which are considered pre-treatment years. This number of pre-treatment outcomes is less than ideal[77]. With this caveat in mind, in what follows, I discuss the specific steps I take to estimate the synthetic control for each impacted municipality.

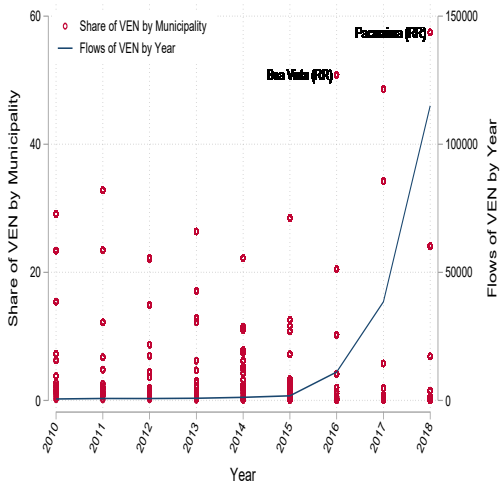
First I need to select the donor pool. My data has 5309 municipalities that did not receive any Venezuelan immigrants after 2017.[78] Following Abadie (2021), I impose the following restrictions on my donor pool. First, and to make sure that the assumption of no interference is held, I discard from the sample all municipalities that received immigrants from any country other than Venezuela. Second, I choose from the donor pool the municipalities that are closest in observable characteristics to the treated municipality. I do this in two steps. Using a lasso estimation on a large set of baseline (2010) variables allows me to choose baseline municipality characteristics that predict the voting outcomes the most. I then use those variables to match, using nearest neighbour matching, each treated municipality to it's closest 100 municipalities from the donor pool. Second, I perform different specifications using different sets of predictor variables and report the results in table III.54. Although adding predictors to the pre-outcome variables results in the lowest RSMPE, I follow the recommendation of Ferman et al. (2018) and use the pre-treatment outcomes as predictors. This approach is transparent and allows for comparability across the different estimations.

[77]I tried to estimate using the first and second round votes as separate points in time, which gives me 10 pre-treatment years. But this approach is not ideal because the elections, although part of same presidential elections, are essentially different in terms of the choices afforded to the voter.

[78]There are municipalities in states other than that of Roraima that report receiving Venezuelans. I exclude those municipalities from my analysis for the following two reasons: Venezuelans represent a miniscule proportion of the total population in these municipalities, and the majority of such municipalities are entry points of immigrants.

Figure III.20: Distribution of Venezuelan Immigrants and Main Municipalities of Entry

(a) Distribution for VEN migrants



(b) Map of Pacaraima and Boa Vista

Figure III.20(a) depicts the share of flows of Venezuelan immigrants as a share municipality population per year and the total number of Venezuelan immigrant flow per year. Figure III.20(b) depicts the significance of the spatial locations of both of Boa Vista and Pacaraima that lie on a major road connecting Venezuela and Brazil. Source: Author's Analysis.

III.F.1 Methodology

Third, I need to estimate the treatment effect. The treatment effect can be described as follows:

$$\alpha_{m,t} = E_{m,t} - \widehat{E^{C}_{m,t}} \tag{III.F.21}$$

where $\widehat{E^{C}_{m,t}}$ is the estimated weighted average of the non-affected donor municipalities.

In order to conduct inference on my treatment estimates, I follow Abadie et al. (2010) and rely on permutation tests. What this does essentially is measure the voting outcomes of municipalities with and without immigrants and compare them. If the outcomes of the treated municipalities are substantially larger or smaller than those calculated using the donor municipalities, then I can conclude with more confidence that the Venezuelan immigration had an impact on the native's voting outcomes. Specifically, I calculate for each municipality m, $\alpha_{m,t}$ as per equation (III.F.21). Then I estimate the root mean square prediction error for each pre-treatment unit as such:

$$RMSPE_m = \sqrt{\frac{\sum_{t=0}^{T-1}(E_{m,t} - \widehat{E^C_{m,t}})}{T-1}} \tag{III.F.22}$$

I then calculate the probability of estimating a result as different as the treated municipality if I were to randomly assign treatment to municipalities in the donor pool as:

$$PValue_m = \sum_{2}^{M} I(|t_m| \geq |t^c_m|)/M, \tag{III.F.23}$$

where M is the total number of municipalities. I follow Ferman and Pinto (2019) and drop from the placebo estimates municipalities with poor pre-intervention fit in order to reduce bias.

Table III.54: Synthetic Pacaraima Root Mean Squared Prediction Error

Panel A: Outcome Lags		Panel B: Predictor Variables	
Specification	RMSPE	Specification	RMSPE
1998-2014 all	.07407	1998-2010 all + 7 predictors	.0662534
1998	.08137	1998-2010 all + 5 predictors	.0657639
1998 & 2002	.08066	1998-2010 all + 4 predictors	.065558
1998 & 2002 & 2006	.0917	1998-2010 all + 3 predictors	.06508
1998 & 2002 & 2006 & 2010	.06958	7 predictors	.16345
2 year averages & 2014	.070879		
1998-2014 average	.069369735		

Notes:This table shows the RMSPE of different specifications in the calculation of the synthetic control estimation. The predictor variables, other than the lagged outcome variables, include municipality level characteristics such as average university level people, poverty rate, per capita GDP and people with an income less than half the minimum wage.

III.F.2 Results

This section shows the results from the synthetic control estimation of the impact of Venezuelan immigration on electoral outcomes. I perform the synthetic control method for the two most populated municipalities in terms of Venezuelan immigrants, Pacaraima and Boa Vista, as well as for each municipality that received Venezuelan im-

Table III.55: Weights and Predictor Means

Donor and Predictor Weights

Municipality Donor Weights		Predictor Weights	
Machado	0.02	Turnout Ratio (2002)	0.25997
Montividiu	0.02	Turnout Ratio (2006)	0.03758
Duas Barras	0.1	Turnout Ratio (2010)	0.11221
Paraguaçu Paulista	0.22	Turnout Ratio (2014)	0.59024
Dores de Campos	0.47		
Pradópolis	0.17		

Actual and Synthetic Pacaraima Predictor Means

	Treated	Synthetic
Turnout Ratio (2002)	78.99	77.71
Turnout Ratio (2006)	80.32	79.03
Turnout Ratio (2010)	81.94	80.68
Turnout Ratio (2014)	84.97	83.82

Notes: This table shows the average value of the predictor variables between the actual and synthetic control unit. This table shows the names, IBGE codes and their respective weights of the donor municipalities involved in creating of the synthetic control unit. This table shows the predictor variables and their respective weights in creating of the synthetic control unit.

migrants after 2017 in the state of Roraima. I present the results for each of Pacariama and Boa Vista, and an estimate of the average of the impact for all the other municipalities. I interpret the latter to represent an approximate estimate of a distributional impact of the Venezuelan immigration across the receiving municipalities.

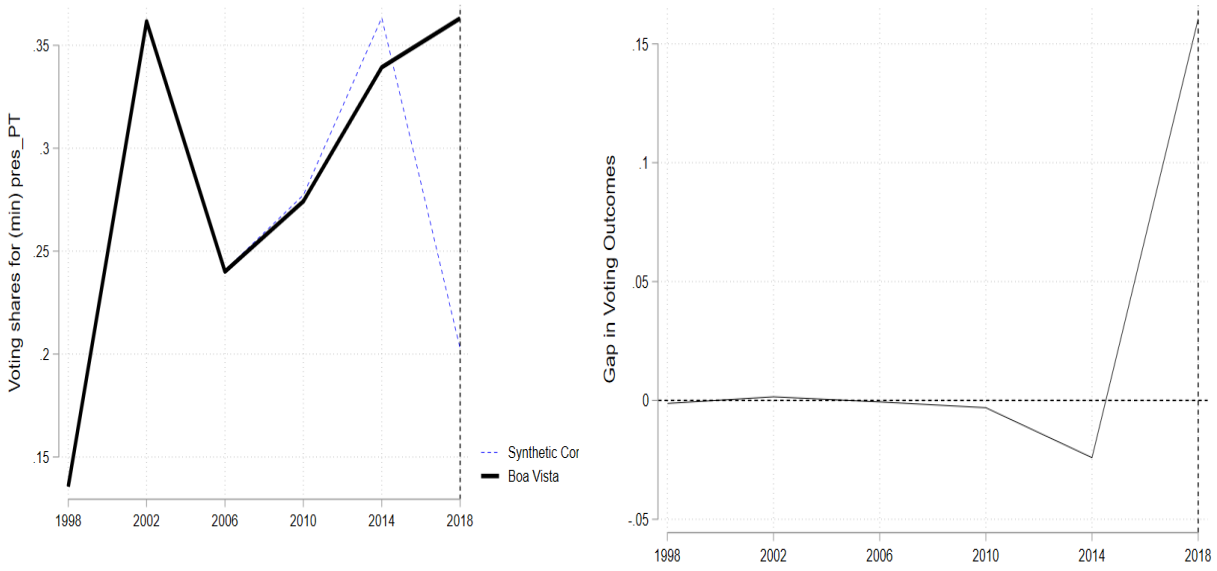
Results are shown for the PT party (left wing presidential party) and the turnout ratio. Other variables are not shown because the synthetic control strategy demands that the variable is available for pre-treatment years. For example, President Bolsonaro's party ran for elections only twice in 2002 (for first round election) and in 2018. The same applies for heterogeneity in the turnout ratio as those are also only available for 2014 and 2018.

My analysis on the impact of Venezuelan immigration finds a significant increase in vote shares towards the left-wing PT party and on turnout ratio, in the municipality of Pacaraima. Upon studying the distribution impacts of Venezuelan immigrants, I find, in contrast, a negative impact on vote shares for the PT party, and a positive impact on the turnout ratio. The increase in vote shares for the PT party holds upon the exclusion of Pacaraima municipality from the analysis.

Figure III.21: Synthetic Control Fit

(a) Synthetic Pacaraima Voting shares for PT party

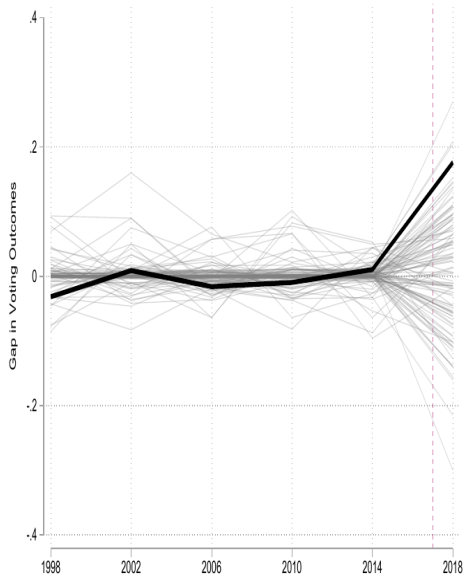
(b) Voting shares for PT party Difference between Pacaraima and Donor States



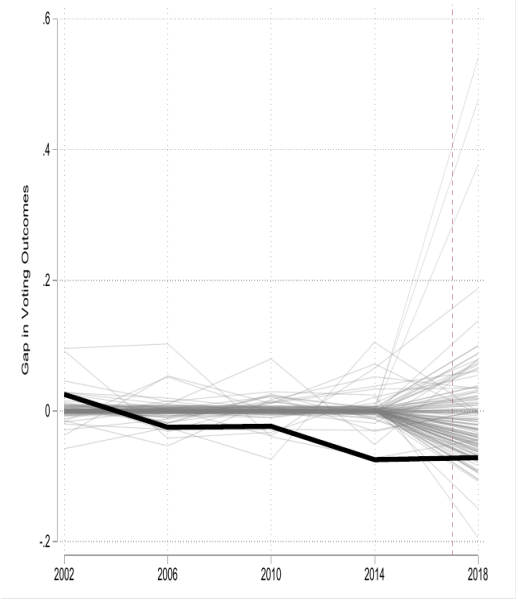
Notes: The figures show the fit of the synthetic control unit to Boa Vista and actual Boa Vista and the gap in voting outcomes between the synthetic and actual. The electoral outcomes represent vote shares(%) per party of the presidential elections. The migration variable is adjusted by population in 2010.

Figure III.22: In-Place Treatment Effects

(a) Difference between Synthetic Pacaraima vs Actual and other placebo states



(b) Difference between Synthetic Boa Vista vs Actual and other placebo states



Notes: The figures show the in-place treatment effects for each municipality in the sample. Solid lines represent the treatment effects for Pacaraima and Boa Vista respectively. The lines represent the vertical distances between the solid and dashed lines displayed in Figure III.21(a) the gap in voting outcomes between the synthetic and actual. The electoral outcomes represent vote shares(%) per party of the presidential elections. The migration variable is adjusted by population in 2010.

Table III.56: Synthetic Control: Distribution Effects

Presidential Elections: First Round		
	Presidential PT Party	Turnout Ratio
Pacaraima	0.134 P-value=0.038	12.42 P-value=.015
Boa Vista	-.0714 P-value= .832	2.557 P-value=0.516
Distribution Impacts of VEN Migrants	-.0156 P-value=0.013	.704 P-value=.0654
Distribution Impacts of VEN Migrants without Pacaraima	-.016 P-value=0.006	.617 P-value=.1132
Distribution Impacts of VEN Migrants without Boa Vista	-.0157 P-value=.0126	0.678 P-value=0.078

The table shows the point estimates and calculated p-values from the Synthetic Control strategy. The p-value departs from its traditional definition and it represents here the probability of estimating a result as extreme as the treated unit's result if the treated unit was randomly assigned among the donor pool. In particular, the pvalue is as estimated in equation (III.F.23). The table shows the results for Pacaraima, Boa Vista and the average estimate of other municipalities (with and without Pacaraima and Boa Vista). Results are for the Turnout Ratio and PT party (left wing) and represents the ratio of votes going to PT party as a proportion of votes to all presidential elections. Both of these variables are as defined in section .

III.F.3 Haitian Entry Points and Comparison with that of Venezuelans

In this section, I replicate the synthetic control analysis for the main entry municipalities of Haitians in 2010. As described in section 3, there were a few number of municipalities, Tabatinga (Amazonas state), Brasiléia (Acre state) or Rio Branco (Acre state), that constituted main entry points for Haitians in the early years of the crisis. After 2014 and due to progress on visa status, those municipalities lost their prominence as the main entry points, and Haitians arrived to Brazil in large numbers and were dispersed across the country. They also used other means of transport, such as airtravel (figure III.17(a)). I do not find that there is any significant impact on turnout ratio, or vote shares to the PT or PSDB parties in municipalities that were initially entry points for Haitians, although the sign of impact on vote shares towards Alckmin is positive. However, I find that there is an increase in turnout ratio and vote share of the PT party in Pacaraima, a main municipality of entry for the most recent Venezuelan immigration.

Table III.57: Synthetic Control: Comparing Results of Haitian versus Venezuelan Entry Municipalities

	Turnout Ratio	PT Party	PSDB Party
Pacaraima (VEN Entry)	12.42 P-value=0.015	0.134 P-value=0.038	0.0017 P-value=0.51
Boa Vista (VEN Entry)	2.557 P-value=0.516	-0.0714 P-value=0.832	0.00045 P-value=0.31
Brasiléia (HTI Entry)	0.025 P-value=0.98	0.00019 P-value=0.64	0.0003 P-value= 0.73
Tabatinga (HTI Entry)	8.97 P-value= 0.55	0.0078 P-value= 0.32	0.00002 P-value= 0.47
Rio Branco (HTI Entry)	2.709 P-value= 0.74	0.0002 P-value= 0.92	0.0001 P-value= 0.75

The table shows the point estimates and calculated p-values from the Synthetic Control strategy applied to five entry points for Haitians and Venezuelans. Those municipalities are Pacaraima and Boa Vista for Venezuelans, and Brasiléia , Tabatinga and Rio Branco for Haitians. Haitian entry points studied here were reported to be one of the main ones in early phases of the refugee inflow, and where refugee camps were initially established. However, in later years and as the flow of Haitians increased, Haitians entered Brazil across different municipalities and used flight as a means of transport. The p-value departs from its traditional definition and it represents here the probability of estimating a result as extreme as the treated unit's result if the treated unit was randomly assigned among the donor pool. In particular, the pvalue is as estimated in equation (III.F.23). Results are for the Turnout Ratio, PT party (left wing), PSDB party (right wing). Both of these variables are as defined in section .

Conclusion

This thesis presents three empirical and quantitative self-contained chapters. The first chapter analyses the impact that the COVID-19 pandemic had on the Brazilian labour markets and discusses the impacts and effectiveness of government policy with a special focus on the generous cash transfer Auxilio Emergencial. The paper highlights the unequal impact that the pandemic had, adversely affecting women, youth, informal workers and workers employed in contact intensive sectors, relative to others. I also find that the generous cash transfer program was successful in cushioning the blow of the pandemic on inequality and poverty. I also presented different counterfactual scenarios on the administration of the cash transfer to minimize leakage. The second and third chapters analyse the impacts of immigration on the host country. Chapter 1 presents evidence on the impact that immigration from 2010 to 2017 had on trade. In this chapter, I study the impact at different levels of product heterogeneity in order to provide evidence of the mechanisms: whether immigrants are affecting the preferences of natives or whether they are able to reduce fixed costs associated with trade. The third chapter, analyses the impact that the Haitian forced immigration had on political behaviour in the 2018 elections. The analysis finds that natives responded to the Haitian presence in their municipality by reducing their voter participation and by casting more votes to the centre-wing party.

The thesis also opens the door to many questions left unanswered that could be promising avenues for research. Contextual research that analyses the impacts of crises, such as the COVID-19 global pandemic, should also take into account the impact that such unexpected crises have on immigrants. Immigrants, whether or not they are fully assimilated into a community (e.g., in terms of their rights to work, education, health or even voting), might find themselves impacted differently than citizens by unforeseen crises due to the type of visa status they hold, the reach of governmental social support programs, and the political system in place.

My work makes use of Brazil's rich data ecosystem to show that a common language can be a predictor of immigrant locations. However, language may not be enough on its own, and therefore research studying migration location choices can include pull factors related to new immigrants sharing a common culture with historic settlements. This common culture can be represented by language, combined with other key characteristics such as religion, level of development, colonial history, among others.

The analysis of the impact of immigrants on trade can go beyond the study of the impact on goods to include the trade of services. The study in the Brazilian context can also expand the range of years studied to include the most recent ones, to encompass the most recent Venezuelan immigration wave. The paper indicated that the trade creation effect of immigrants differs based on the country of origin. However, this is far from being an exhaustive list and future work can include other characteristics that vary by the country of origin, and ones that are at an individual level and differ based on the immigrant's level of education or type of profession. Studying the impacts of trade at the firm level based on migrant employees is yet another important avenue that can allow for a better understanding of the channels at play. The analysis of the impact of the Haitian forced immigration on political behaviour can be expanded in three important ways. For one, the study can be expanded to include the impact from the Venezuelan immigration and to compare this impact with that of the Haitians. Expanding the analysis can also incorporate the 2022 presidential election where Jair Bolsonaro will be competing with the well-known and the creator of Bolsa Família Lula da Silva. Second, the study can delve deeper into the labour market outcomes within the formal sector to separately identify, at a sectoral level, if the Haitian immigrant represented a substitute or a complement to the native worker. And third, this analysis, can go beyond the political outcomes from the elections, to study how these changes in electoral outcomes affect the social, political and economic policies.

Each one of the chapters presented in this thesis is situated within the Brazilian context, and hence there is some commonality between the chapters. However, the findings need not be seen as entirely specific to Brazil. The analysis might well be applicable in other middle-income countries. Chapter one informs policy in two ways. First, depending on ex-ante susceptibility, individuals may be impacted by crises differently and therefore a tailored approach to mitigate the impact on those who are most affected could be more effective than blanket approaches – and potentially fiscally more sustainable. Second, governmental cash transfers can be very effective in cushioning the impact of crises on short-term poverty and inequality. The success of the Brazilian

cash transfer Auxílio Emergencial rested in its speed to reach beneficiaries, which was made possible thanks to the administrative datasets available through Bolsa Família. However, the program came at a large fiscal cost and therefore consideration should be given to finance additional social spending within the expenditure limits imposed by the fiscal ceiling. Chapter two shows that immigrants have a positive impact on both exports and imports. However, it has been documented that immigrants may pose a strain on government expenditure and affect the wages and employment opportunities of native workers with substitutable skills to that of the immigrants. Therefore, there are trade-offs that need to be considered to inform potential distribution policies for the receiving country. My third chapter shows that an episode of forced immigration, particularly one from a country of origin that doesn't traditionally send immigrants, can have important implications on voter behaviour. In particular, a key impact to consider is the impact that immigration has on political turnout. Even for an immigration wave that doesn't exceed the 1% from total population in the host country, the adverse impacts on voter participation can pose a threat to representation in a democratic system. The roots behind such a finding lie in cultural attitudes towards immigrants and in voter uncertainty between candidates. Therefore, state interventions or non-governmental targeted campaigns that tackle these underlying channels can help reduce voter absenteeism. Further, the paper presents evidence that areas that received Haitian migrants also saw positive economic outcomes relative to those that didn't. Contingent on the relative size of the immigration wave to the population and the status of the economy, my finding indicates that there is potential to consider inclusive policies that promote labour market assimilation of migrants in receiving countries, similar to the laws that Brazil implemented.

Bibliography

Abid, R. Z., Manan, S. A., and Rahman, Z. A. A. A. (2017). 'a flood of syrians has slowed to a trickle': The use of metaphors in the representation of syrian refugees in the online media news reports of host and non-host countries. *Discourse & Communication*, 11(2):121–140.

Adams, J., Dow, J., and Merrill, S. (2006). The political consequences of alienation-based and indifference-based voter abstention: Applications to presidential elections. *Political Behavior*, 28(1):65–86.

Adams, J. and Merrill III, S. (2003). Voter turnout and candidate strategies in american elections. *The Journal of Politics*, 65(1):161–189.

Adams-Prassl, A., Boneva, T., Golin, M., and Rauh, C. (2020a). Inequality in the impact of the coronavirus shock: Evidence from real time surveys. *Journal of Public economics*, 189:104245.

Adams-Prassl, A., Boneva, T., Golin, M., and Rauh, C. (2020b). Inequality in the impact of the coronavirus shock: new survey evidence for the uk.

Adams-Prassl, A., Boneva, T., Golin, M., and Rauh, C. (2020c). Work tasks that can be done from home: Evidence on variation within & across occupations and industries.

Akarca, A. T. and Tansel, A. (2015). Impact of internal migration on political participation in turkey. *IZA Journal of Migration*, 4(1):1–14.

Al Masri, D., Flamini, V., and Toscani, F. (2021). The short-term impact of covid-19 on labor markets, poverty and inequality in brazil.

Al Masri, D. and Vargas-Silva, C. (2020). The impact of migration on trade: Evidence from brazil. *Inter-American Development Bank*.

Alesina, A., Devleeschauwer, A., Easterly, W., Kurlat, S., and Wacziarg, R. (2003). Fractionalization. *Journal of Economic growth*, 8(2):155–194.

Alesina, A., Miano, A., and Stantcheva, S. (2018). Immigration and redistribution. Technical report, National Bureau of Economic Research.

Alfaro, L., Becerra, O., and Eslava, M. (2020). Emes and COVID-19: shutting down in a world of informal and tiny firms. Technical report, National Bureau of Economic Research.

Allport, G. W., Clark, K., and Pettigrew, T. (1954). The nature of prejudice.

Alon, T., Doepke, M., Olmstead-Rumsey, J., and Tertilt, M. (2020). The impact of COVID-19 on gender equality. Technical report, National Bureau of economic research.

Altonji, J. G. and Card, D. (1991). The effects of immigration on the labor market outcomes of less-skilled natives. In *Immigration, trade, and the labor market*, pages 201–234. University of Chicago Press.

Alves, I. P. (2018). O fluxo migratório venezuelano para o brazil como uma questão amazônica. *Migrações Venezuelanas. Universidade Estadual de Campinas*.

Anatolyev, S. (2013). Instrumental variables estimation and inference in the presence of many exogenous regressors. *The Econometrics Journal*, 16(1):27–72.

Angrist, J. D., Imbens, G. W., and Krueger, A. B. (1999). Jackknife instrumental variables estimation. *Journal of Applied Econometrics*, 14(1):57–67.

Angrist, J. D., Imbens, G. W., and Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American statistical Association*, 91(434):444–455.

Apedo-Amah, M. C., Avdiu, B., Cirera, X., Cruz, M., Davies, E., Grover, A., Iacovone, L., Kilinc, U., Medvedev, D., Maduko, F. O., et al. (2020). Unmasking the impact of COVID-19 on businesses.

Arbatlı, C. E., Ashraf, Q. H., Galor, O., and Klemp, M. (2020). Diversity and conflict. *Econometrica*, 88(2):727–797.

Armingeon, K. and Schädel, L. (2015). Social inequality in political participation: The dark sides of individualisation. *West European Politics*, 38(1):1–27.

Arrow, K. J. (1972). Gifts and exchanges. *Philosophy & Public Affairs*, pages 343–362.

Assunção, J., Verdier, T., and de Brito, C. H. G. (2020). Immigrants' pro-trade effects: Evidence from brazil.

Audebert, C. (2011). Cnrs, migrinter (umr 6588, université de poitiers). *Le défi haïtien*, page 193.

Aydemir, A. B. and Kırdar, M. (2014). Integration of 1989 repatriates from bulgaria to the turkish labor market.

Azevedo, A. M. and Dantas, A. C. P. (2017). Inadequate integration of refugees in brazil as a violation to the right of human dignity. *Revista do CEPEJ*, (20).

Baker, A. and Greene, K. F. Latin American Election Results with Party Ideology Scores. 2.0 (1993 to present). https://www.dropbox.com/sh/lu5pteaoqynx15j/AACOuOvmcvDuigqaHp5W1ThLa?dl=0.

Baker, A. and Greene, K. F. (2011). The latin american left's mandate: free-market policies and issue voting in new democracies. *World Politics*, 63(1):43–77.

Baltagi, B. H., Egger, P. H., and Pfaffermayr, M. (2014). Panel data gravity models of international trade. *Available at SSRN 2398292*.

Bandyopadhyay, S., Coughlin, C. C., and Wall, H. J. (2008). Ethnic networks and us exports. *Review of International Economics*, 16(1):199–213.

Baqaee, D. and Farhi, E. (2022). Supply and demand in disaggregated keynesian economies with an application to the COVID-19 crisis. *American Economic Review*, 112(5):1397–1436.

Barber, M. and Imai, K. (2014). Estimating neighborhood effects on turnout from geocoded voter registration records. *Princeton, NJ: Mimeo*.

Barberia, L. G. and Piazza, K. S. (2021). The auxílio emergencial and stringent social distancing policies in brazil: The impact on mobility and non-work travel outside the home. *Revista Latinoamericana de Opinión Pública*, 10(2):93–135.

Bardhan, A. D. and Guhathakurta, S. (2004). Global linkages of subnational regions: coastal exports and international networks. *Contemporary Economic Policy*, 22(2):225–236.

Barrot, J.-N., Grassi, B., and Sauvagnat, J. (2021). Sectoral effects of social distancing. In *AEA Papers and Proceedings*, volume 111, pages 277–81.

Bartik, A. W., Bertrand, M., Cullen, Z., Glaeser, E. L., Luca, M., and Stanton, C. (2020a). The impact of COVID-19 on small business outcomes and expectations. *Proceedings of the national academy of sciences*, 117(30):17656–17666.

Bartik, A. W., Cullen, Z. B., Glaeser, E. L., Luca, M., and Stanton, C. T. (2020b). What jobs are being done at home during the COVID-19 crisis? evidence from firm-level surveys. Technical report, National Bureau of Economic Research.

Battisti, M., Felbermayr, G., Peri, G., and Poutvaara, P. (2018). Immigration, search and redistribution: A quantitative assessment of native welfare. *Journal of the European Economic Association*, 16(4):1137–1188.

Bauer, T., Epstein, G. S., and Gang, I. N. (2005). Enclaves, language, and the location choice of migrants. *Journal of population Economics*, 18(4):649–662.

Beekma, I. (2015). Realities of the promised land: Life for haitian immigrants in brazil is not always as they imagined–especially as some employers deny them rights. *Index on Censorship*, 44(1):34–37.

Bellettini, G., Ceroni, C. B., and Monfardini, C. (2016). Neighborhood heterogeneity and electoral turnout. *Electoral Studies*, 42:146–156.

Benhabib, J. (1996). On the political economy of immigration. *European Economic Review*, 40(9):1737–1743.

Benzeval, M., Burton, J., Crossley, T. F., Fisher, P., Jäckle, A., Low, H., and Read, B. (2020). The idiosyncratic impact of an aggregate shock: the distributional consequences of COVID-19. *Available at SSRN 3615691*.

Best, R. and Burke, P. J. (2019). Macroeconomic impacts of the 2010 earthquake in haiti. *Empirical Economics*, 56(5):1647–1681.

Borjas, G. J. (1999). The economic analysis of immigration. *Handbook of labor economics*, 3:1697–1760.

Borjas, G. J. (2014). *Immigration economics*. Harvard University Press.

Borri, N., Drago, F., Santantonio, C., and Sobbrio, F. (2021). The "great lockdown": Inactive workers and mortality by COVID-19. *Health economics*, 30(10):2367–2382.

Borusyak, K., Hull, P., and Jaravel, X. (2022). Quasi-experimental shift-share research designs. *The Review of Economic Studies*, 89(1):181–213.

Bound, J., Jaeger, D. A., and Baker, R. M. (1995). Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. *Journal of the American statistical association*, 90(430):443–450.

Bouvier, M., Roubaud, F., Razafindrakoto, M., Teixeira, R., et al. (2022). Labour market transitions in the time of COVID-19 in brazil: a panel data analysis. Technical report.

Braga, C. C. V. (2010). Minustah and the security environment in haiti: Brazil and south american cooperation in the field. *International Peacekeeping*, 17(5):711–722.

Bredtmann, J., Nowotny, K., and Otten, S. (2020). Linguistic distance, networks and migrants' regional location choice. *Labour Economics*, 65:101863.

Briant, A., Combes, P.-P., and Lafourcade, M. (2014). Product complexity, quality of institutions and the protrade effect of immigrants. *The World Economy*, 37(1):63–85.

Broda, C., Greenfield, J., and Weinstein, D. (2006). From groundnuts to globalization: A structural estimate of trade and growth.

Broda, C. and Weinstein, D. E. (2006). Globalization and the gains from variety. *The Quarterly journal of economics*, 121(2):541–585.

Bulawayo, M., Mudenda, D., Ndulo, M., and Simwanza, C. (2020). Does immigration stimulate non-traditional exports? evidence from zambia. *African Journal of Economic Review*, 8(3):116–136.

Burger, M., Van Oort, F., and Linders, G.-J. (2009). On the specification of the gravity model of trade: zeros, excess zeros and zero-inflated estimation. *Spatial economic analysis*, 4(2):167–190.

Burgoon, B. and Rooduijn, M. (2021). 'immigrationization' of welfare politics? anti-immigration and welfare attitudes in context. *West European Politics*, 44(2):177–203.

Cajner, T., Crane, L. D., Decker, R., Hamins-Puertolas, A., and Kurz, C. J. (2020). Tracking labor market developments during the COVID-19 pandemic: A preliminary assessment.

Calonico, S., Cattaneo, M. D., and Farrell, M. H. (2018). On the effect of bias estimation on coverage accuracy in nonparametric inference. *Journal of the American Statistical Association*, 113(522):767–779.

Calonico, S., Cattaneo, M. D., and Titiunik, R. (2014). Robust data-driven inference in the regression-discontinuity design. *The Stata Journal*, 14(4):909–946.

Campbell, D. E. (2013). Social networks and political participation. *Annual Review of Political Science*, 16:33–48.

Cantor, D. J., Freier, L. F., and Gauci, J.-P. (2015). *A liberal tide?: Immigration and asylum law and policy in Latin America*. Institute of Latin American Studies London.

Capron, H. and Kruseman, J.-L. (1988). Is political rivalry an incentive to vote? *Public Choice*, pages 31–43.

Card, D. (2001). Immigrant inflows, native outflows, and the local labor market impacts of higher immigration. *Journal of Labor Economics*, 19(1):22–64.

Card, D., Dustmann, C., and Preston, I. (2012). Immigration, wages, and compositional amenities. *Journal of the European Economic Association*, 10(1):78–119.

Card, D. and Peri, G. (2016). Immigration economics by george j. borjas: a review essay. *Journal of Economic Literature*, 54(4):1333–49.

Carvalho, C. (2012a). Acre sofre com invasão de imigrantes do haiti. *O Globo*, 1.

Carvalho, C. (2012b). Acre sofre com invasão de imigrantes do haiti. *O Globo*, 1.

Carvalho, C. (2012c). Peru e bolívia, vizinhos do brasil, negam ajuda a haitianos. *O Globo*, 8.

Carvalho, C. (2012d). Peru e bolívia, vizinhos do brasil, negam ajuda a haitianos. *O Globo*, 8.

Cattaneo, M. D., Jansson, M., and Ma, X. (2020). Simple local polynomial density estimators. *Journal of the American Statistical Association*, 115(531):1449–1455.

Cavalcanti, L., Oliveira, A. T., Tonhati, T., and Dutra, D. (2015). A inserção dos imigrantes no mercado de trabalho brasileiro. *Relatório anual*.

Cavalcanti, L., Tonhati, T., Araújo, D., Brasil, E., and Oliveira, T. (2017). Haitian immigrants in the brazilian labour market: an overview. *Rev Est Pesq Am*, 11(1):192–203.

Cavallo, E., Powell, A., and Becerra, O. (2010). Estimating the direct economic damages of the earthquake in haiti. *The Economic Journal*, 120(546):F298–F312.

Centro De Direitos Humanos E Cidadania Do Imigrante/CDHIC (2013). Organizações lançam manifesto pela humanização da cobertura midiática sobre as migrações no brazil. *CDHIC*.

Chaney, T. (2008). Distorted gravity: heterogeneous firms, market structure, and the geography of international trade. *American Economic Review*, 98(4):1707–1721.

Chepkemoi, J. (2019). What languages are spoken in haiti?

Clemens, M. A., Lewis, E. G., and Postel, H. M. (2018). Immigration restrictions as active labor market policy: Evidence from the mexican bracero exclusion. *American Economic Review*, 108(6):1468–87.

Clogg, C. C., Petkova, E., and Haritou, A. (1995). Statistical methods for comparing regression coefficients between models. *American journal of sociology*, 100(5):1261–1293.

Co, C. Y., Euzent, P., and Martin, T. (2004). The export effect of immigration into the usa. *Applied Economics*, 36(6):573–583.

Coady, M. D. and Le, N.-P. (2020). *Designing fiscal redistribution: the role of universal and targeted transfers*. International Monetary Fund.

Cogo, D. (2013). *Guia das migrações transnacionais e diversidade cultural para comunicadores: migrantes no Brasil*. INCOM UAB.

Cogo, D. (2014). Haitianos no brasil. comunicação e interação em redes migratórias transnacionais. *Chasqui. Revista Latinoamericana de Comunicacion*, (125):23–32.

Cogo, D. and Silva, T. (2016). Entre a" fuga" e a" invasão": alteridade e cidadania da imigração haitiana na mídia brasileira. *Revista Famecos*, 23(1):ID21885–ID21885.

Coibion, O., Gorodnichenko, Y., and Weber, M. (2020). Labor markets during the COVID-19 crisis: A preliminary view. Technical report, National Bureau of economic research.

Combes, P.-P., Lafourcade, M., and Mayer, T. (2005). The trade-creating effects of business and social networks: evidence from france. *Journal of international Economics*, 66(1):1–29.

Coppedge, M. (1998). The dynamic diversity of latin american party systems. *Party Politics*, 4(4):547–568.

Crosby, D. A., Dowsett, C. J., Gennetian, L. A., and Huston, A. C. (2010). A tale of two methods: comparing regression and instrumental variables estimates of the effects of preschool child care type on the subsequent externalizing behavior of children in low-income families. *Developmental psychology*, 46(5):1030.

da Frota Simões, G. (2017). Perfil sociodemográfico e laboral da imigração venezuelana no brasil. *Curitiba, BR: CRV*.

Dabène, O. and Louault, F. (2018). *Atlas du Brésil. Promesses et défis d'une puissance émergente*. Autrement.

Dalton, R. J. and Klingemann, H.-D. (2009). Overview of political behavior. In *The Oxford handbook of political science*.

Damé, L. (2013). Governo fará força tarefa para regularizar situação dos haitianos no acre. *O Globo, Brasília*, 10.

de Andreazzi, C. S., Brandão, M. L., Bueno, M. G., Winck, G. R., Rocha, F. L., Raimundo, R. L., Metzger, J. P., Chame, M., Cordeiro, J. L. P., and D'Andrea, P. S. (2020). Brazil's COVID-19 response. *The Lancet*, 396(10254):e30.

de Lima, P. A., Brisola, E. M. A., and da Silva, A. L. (2019). Tendências antidemocráricas em partidos políticos de um município paulista. *Interação-Revista de Ensino, Pesquisa e Extensão*, 21(1):92–108.

de Souza, A. R. and da Silveira, M. D. C. P. (2018). O fluxo migratório de venezuelanos para o brasil (2014-2018): análise do arcabouço jurídico brasileiro e da conjuntura interna venezuelana. *Brazilian Journal of Latin American Studies*, 17(32):114–132.

Debandi, N., Patallo, M. F., Leguizamón, C., Argentieri, C., Gurrieri, J., Texido, E., Modolo, V., and Miranda, M. E. (2017). Diagnóstico regional sobre migración haitiana.

Del Vecchio, V. A. and Almeida, V. B. F. (2018). Panorama do fluxo migratório de venezuelanos no brasil e américa latina. *BAENINGER, Rosana; SILVA, João Carlos Jarochinski (coordenadores). Migrações Venezuelanas. Campinas, SP: Núcleo de Estudos de População" Elza Berquó"–NEPO/UNICAMP*, pages 158–163.

Denny, E. K. (2017). *Poverty and the Psychology of Political (In) action*. University of California, San Diego.

Dingel, J. and Neiman, B. (2020). How many jobs can be done at home?, nber working paper.

Downs, A. et al. (1957). An economic theory of democracy.

Dunlevy, J. A. (2006). The influence of corruption and language on the protrade effect of immigrants: Evidence from the american states. *Review of Economics and Statistics*, 88(1):182–186.

Dustmann, C., Schönberg, U., and Stuhler, J. (2017). Labor supply shocks, native wages, and the adjustment of local employment. *The Quarterly Journal of Economics*, 132(1):435–483.

Dustmann, C., Vasiljeva, K., and Piil Damm, A. (2019). Refugee migration and electoral outcomes. *The Review of Economic Studies*, 86(5):2035–2091.

e Silva, L. V., de Andrade Abi, M. D. P., Dos Santos, A. M. T. B., de Mattos Teixeira, C. A., Gomes, V. H. M., Cardoso, E. H. S., da Silva, M. S., Vijaykumar, N., Carvalho, S. V., Frances, C. R. L., et al. (2020). COVID-19 mortality underreporting in brazil: analysis of data from government internet portals. *Journal of medical Internet research*, 22(8):e21413.

Ehrhart, H., Le Goff, M., Rocher, E., and Singh, R. J. (2014). Does migration foster exports? evidence from africa. *Evidence from Africa (January 1, 2014). World Bank Policy Research Working Paper*, (6739).

Esnard, A.-M. and Sapat, A. (2011). Disasters, diasporas and host communities: Insights in the aftermath of the haiti earthquake. *Journal of Disaster Research*, 6(3).

Espinoza, M. V. and Brumat, L. (2018). Brazil elections 2018: how will bolsonaro's victory affect migration policy in brazil and south america? *LSE Latin America and Caribbean Blog*.

Esses, V. M., Medianu, S., and Lawson, A. S. (2013). Uncertainty, threat, and the role of the media in promoting the dehumanization of immigrants and refugees. *Journal of social issues*, 69(3):518–536.

Estarque, M. (2014). No centro de disputa política, haitianos vivem incerteza em são paulo. *DW Brasil, São Paulo*, 30.

Facchini, G. and Mayda, A. M. (2009). Does the welfare state affect individual attitudes toward immigrants? evidence across countries. *The review of economics and statistics*, 91(2):295–314.

Farias, E. (2011). Estado brasileiro nunca deu apoio a imigrantes haitianos, diz voluntária do amazonas. *A crítica. Manaus*, 24.

Felbermayr, G., Grossmann, V., and Kohler, W. (2015). Migration, international trade, and capital formation: Cause or effect? In *Handbook of the economics of international migration*, volume 1, pages 913–1025. Elsevier.

Felbermayr, G. J. and Jung, B. (2009). The pro-trade effect of the brain drain: Sorting out confounding factors. *Economics Letters*, 104(2):72–75.

Felbermayr, G. J., Jung, B., and Toubal, F. (2010). Ethnic networks, information, and international trade: Revisiting the evidence. *Annals of Economics and Statistics/Annales d'Économie et de Statistique*, pages 41–70.

Felbermayr, G. J. and Toubal, F. (2012). Revisiting the trade-migration nexus: Evidence from new oecd data. *World Development*, 40(5):928–937.

Fergusson, L., Querubin, P., Ruiz, N. A., and Vargas, J. F. (2021). The real winner's curse. *American Journal of Political Science*, 65(1):52–68.

Ferigato, S., Fernandez, M., Amorim, M., Ambrogi, I., Fernandes, L. M., and Pacheco, R. (2020). The brazilian government's mistakes in responding to the COVID-19 pandemic. *The Lancet*, 396(10263):1636.

Fernandes, D., Milesi, R., and Farias, A. From haiti to brazil: the new migratory flow. *Instituto Migrações e Direitos Humanos (IMDH)*.

Fieldhouse, E. and Cutts, D. (2008). Diversity, density and turnout: The effect of neighbourhood ethno-religious composition on voter turnout in britain. *Political Geography*, 27(5):530–548.

Foged, M. and Peri, G. (2016). Immigrants' effect on native workers: New analysis on longitudinal data. *American Economic Journal: Applied Economics*, 8(2):1–34.

Frayssinet, F. (2012). Diáspora haitiana coloca política imigratória brasileira em xeque. *Site Opera Mundi. Rio de Janeiro*, 17.

Freire, A. and Turgeon, M. (2020). Random votes under compulsory voting: Evidence from brazil. *Electoral Studies*, 66:102168.

Gagnon, J., Khoudour-Castéras, D., et al. (2011). *Tackling the policy challenges of migration*. OECD: Organisation for Economic Co-operation and Development.

Gallego, A. (2007). Unequal political participation in europe. *International Journal of Sociology*, 37(4):10–25.

Garcia, M. L. T., Pandolfi, A. F., Leal, F. X., Stocco, A. F., Borrego, A. E., Borges, R. E., Oliveira, E. F. d. A., Lang, A. E., Andrade, C. O., Salazar, S. N., et al. (2021). The COVID-19 pandemic, emergency aid and social work in brazil. *Qualitative Social Work*, 20(1-2):356–365.

Gedan, B. N. (2017). Venezuelan migration: Is the western hemisphere prepared for a refugee crisis? *SAIS review of International Affairs*, 37(2):57–64.

Genc, M., Gheasi, M., Nijkamp, P., and Poot, J. (2012). The impact of immigration on international trade: a meta-analysis. In *Migration impact assessment*. Edward Elgar Publishing.

Gennetian, L. A., Magnuson, K., and Morris, P. A. (2008). From statistical associations to causation: what developmentalists can learn from instrumental variables techniques coupled with experimental data. *Developmental psychology*, 44(2):381.

Ginsburgh, V. and Weber, S. (2020). The economics of language. *Journal of Economic Literature*, 58(2):348–404.

Girma, S. and Yu, Z. (2002). The link between immigration and trade: Evidence from the united kingdom. *Weltwirtschaftliches Archiv*, 138(1):115–130.

Goldsmith-Pinkham, P., Sorkin, I., and Swift, H. (2020). Bartik instruments: What, when, why, and how. *American Economic Review*, 110(8):2586–2624.

Gould, D. M. (1994). Immigrant links to the home country: empirical implications for us bilateral trade flows. *The Review of Economics and Statistics*, pages 302–316.

Guerrero-Amezaga, M. E., Humphries, J. E., Neilson, C. A., Shimberg, N., and Ulyssea, G. (2022). Small firms and the pandemic: Evidence from latin america. *Journal of Development Economics*, 155:102775.

Hale, T., Angrist, N., Goldszmidt, R., Kira, B., Petherick, A., Phillips, T., Webster, S., Cameron-Blake, E., Hallas, L., Majumdar, S., et al. (2021). A global panel database of pandemic policies (oxford COVID-19 government response tracker). *Nature human behaviour*, 5(4):529–538.

Hausman, J. and McFadden, D. (1984). Specification tests for the multinomial logit model. *Econometrica: Journal of the econometric society*, pages 1219–1240.

Head, K. and Ries, J. (1998). Immigration and trade creation: econometric evidence from canada. *Canadian journal of economics*, pages 47–62.

Hemisphere, I.M.F.W. (2020). *Regional Economic Outlook, October 2020, Western Hemisphere: Pandemic Persistence Clouds the Recovery*. International Monetary Fund.

Herander, M. G. and Saavedra, L. A. (2005). Exports and the structure of immigrant-based networks: the role of geographic proximity. *Review of Economics and Statistics*, 87(2):323–335.

Holland, M. (2019). Fiscal crisis in brazil: causes and remedy. *Brazilian Journal of Political Economy*, 39:88–107.

Humphries, J. E., Neilson, C. A., and Ulyssea, G. (2020). Information frictions and access to the paycheck protection program. *Journal of public economics*, 190:104244.

Imbens, G. W. and Wooldridge, J. M. (2009). Recent developments in the econometrics of program evaluation. *Journal of economic literature*, 47(1):5–86.

Institute for Fiscal Studies (IFS) (2020). COVID-19 and inequalities. *Inequality: The IFS Deaton Review*.

International Monetary Fund (2017). Fiscal monitor: Tackling inequality.

International Organization for Migration. Key facts you should know about global migration trends. *International Organization for Migration*.

International Organization for Migration (IOM). Number of refugees, migrants from venezuela reaches 3 million. *IOM UN Migration*.

International Trade Centre (2020a). *COVID-19: The Great Lockdown and its Impact on Small Business*. International Trade and Investment.

International Trade Centre (2020b). *COVID-19: The Great Lockdown and its Impact on Small Business*. International Trade and Investment.

Jadotte, E. (2012). Brain drain, brain circulation and diaspora networks in haiti. *UNCTAD background paper for The Least Developed Countries Report*.

Jaeger, D. A., Ruist, J., and Stuhler, J. (2018). Shift-share instruments and the impact of immigration. Technical report, National Bureau of Economic Research.

Jaupart, P. (2018). Divided island: Haitian immigration and electoral outcomes in the dominican republic. *Journal of Economic Geography*, 18(4):951–999.

John, M. (2019). Venezuelan economic crisis: crossing latin american and caribbean borders. *Migration and Development*, 8(3):437–447.

Jöst, P. (2021). Where do the less affluent vote? the effect of neighbourhood social context on individual voting intentions in england. *Political Studies*, page 00323217211027480.

Jubilut, L. L., Sombra Muiños de Andrade, C., and de Lima Madureira, A. (2016). Visas humanitaires : s'appuyer sur l'expérience du brésil. *Forced Migration Review (FMR)*, 53.

Kaniovski, S. and Mueller, D. C. (2006). Community size, heterogeneity and voter turnouts. *Public Choice*, 129(3):399–415.

Kern, A., Marien, S., and Hooghe, M. (2015). Economic crisis and levels of political participation in europe (2002–2010): The role of resources and grievances. *West European Politics*, 38(3):465–490.

Klofstad, C. (2010). *Civic talk: Peers, politics, and the future of democracy*. Temple University Press.

Kolesár, M., Chetty, R., Friedman, J., Glaeser, E., and Imbens, G. W. (2015). Identification and inference with many invalid instruments. *Journal of Business & Economic Statistics*, 33(4):474–484.

Komatsuzaki, T., Pienknagura, S., Pizzinelli, C., Roldós, J., and Toscani, F. (2020). Latin american labor markets during COVID-19. *Regional Economic Outlook for Western Hemisphere: Pandemic Persistence Clouds the Recovery*.

König, M. and Winkler, A. (2021). COVID-19: Lockdowns, fatality rates and gdp growth. *Intereconomics*, 56(1):32–39.

Lambert, P., Van Reenen, J., et al. (2021). *A Major Wave of UK Business Closures by April 2021?: The Scale of the Problem and what Can be Done*. Centre for Economic Performance, London School of Economics and Political ....

Lara De Arruda, P., Lazarotto de Andrade, M., Falcão, T., Teixeira Barbosa, D., and Morgandi, M. (2021). Auxílio emergencial-lessons from the brazilian experience responding to COVID-19.

Lara de Arruda, P., Lazarotto de Andrade, M., Falcao, T., Teixeira Barbosa, D., and Morgandi, M. (2022). Challenges and potentialities for implementing social protection responses to emergency through decentralized administration.

Law, W. (2021). Decomposing political advertising effects on vote choices. *Public Choice*, 188:525–547.

Lazear, E. P. (1999). Culture and language. *Journal of political Economy*, 107(S6):S95–S126.

Leer, J., DeAntonio, D., Sweet, R., deRitis, C., and Zandi, M. (2020). Business closures and entrepreneurship during covid-19. *Moody's Analytics Morning Consult*.

Leibovici, F., Santacreu, A. M., and Famiglietti, M. (2020). Social distancing and contact-intensive occupations. *On the economy, St. Louis FED*.

Leyva, G. and Urrutia, C. (2022). Informal labor markets in times of pandemic. *Review of Economic Dynamics*.

Lima, M. C. (2012a). Right wing politics in contemporary brazil. *World Affairs: The Journal of International Issues*, 16(1):40–59.

Lima, W. (2012b). Haitianos refugiados no acre são enviados a sp e porto alegre. *Portal iG Maranhão*, 9.

Lin, K.-H., Aragão, C., and Dominguez, G. (2021). Firm size and employment during the pandemic. *Socius*, 7:2378023121992601.

Llull, J. (2018). Immigration, wages, and education: A labour market equilibrium structural model. *The Review of Economic Studies*, 85(3):1852–1896.

Lodola, G. and Queirolo, R. (2005). Ideological classification of latin american political parties. *Pittsburgh: University of Pittsburgh*.

Londoño, E. and Darlington, S. (2018). Jair bolsonaro wins brazil's presidency, in a shift to the far right. *The New York Times*.

Long, J. S. and Freese, J. (2006). *Regression models for categorical dependent variables using Stata*, volume 7. Stata press.

Longaker, R. P. (1961). E. e. schattschneider. the semi- sovereign people: A realist's view of democracy in america. pp. viii, 147. new york: Holt, rinehart and winston, 1960. $2.95 clothbound; $1.75 paper- bound. *The ANNALS of the American Academy of Political and Social Science*, 338(1):146–147.

Marcel, Y. and Natani, R. (2013). Acre decreta situação de emergência social por causa de surto de imigração. *Portal G1, Acre*, 9.

Marcinkiewicz, K. (2018). The economy or an urban–rural divide? explaining spatial patterns of voting behaviour in poland. *East European Politics and Societies*, 32(04):693–719.

Mayda, A. M. (2006). Who is against immigration? a cross-country investigation of individual attitudes toward immigrants. *The review of Economics and Statistics*, 88(3):510–530.

Mayda, A. M., Peri, G., and Steingress, W. (2022). The political impact of immigration: Evidence from the united states. *American Economic Journal: Applied Economics*, 14(1):358–89.

Mckenzie, D. (2018). Rethinking identification under the bartik shift-share instrument.

Mellon, J., Evans, G., Fieldhouse, E., Green, J., and Prosser, C. (2018). Opening the can of worms: Most existing studies of aggregate level turnout are meaningless. *SSRN's eLibrary*.

Mijares, V. M. and Rojas Silva, N. (2018). Venezuelan migration crisis puts the region's democratic governability at risk.

Millimet, D. L. and Osang, T. (2007). Do state borders matter for us intranational trade? the role of history and internal migration. *Canadian Journal of Economics/Revue canadienne d'économique*, 40(1):93–126.

Ministério Da Justiça E Segurança Pública (2019). Novas resoluções normativas do conselho nacional de imigração-cnig no 125 no 97/12. *Portal de Imigração Ministério Da Justiça E Segurança Pública*.

Miura, H. (2014). The haitian migration flow to brazil: Aftermath of the 2010 earthquake. *The State of Environmental Migration 2014. A Review of 2013*.

Mongey, S., Pilossoph, L., and Weinberg, A. (2021). Which workers bear the burden of social distancing? *The Journal of Economic Inequality*, 19(3):509–526.

Monras, J. (2020). Immigration and wage dynamics: Evidence from the mexican peso crisis. *Journal of Political Economy*, 128(8):3017–3089.

Moreira, J. B. (2015). Política externa e refugiados no brasil: uma análise sobre o governo lula (2003-2010). *Carta Internacional*, 10(3):133–151.

MPI (2010). Natural disasters in haiti and pakistan highlight diaspora response. *Migration Policy Institute*.

Müller, A., Wichmann, S., Velupillai, V., Brown, C. H., Brown, P., Sauppe, S., Holman, E. W., Bakker, D., List, J.-M., Egorov, D., et al. (2010). Asjp world language tree of lexical similarity: Version 3 (july 2010). *Retrieved*, 10(19):2015.

Müller, M.-M. and Steinke, A. (2020). The geopolitics of brazilian peacekeeping and the united nations' turn towards stabilisation in haiti. *Peacebuilding*, 8(1):54–77.

Mutz, D. C. (2006). *Hearing the other side: Deliberative versus participatory democracy*. Cambridge University Press.

Neto, V. P. (2018). Conditional cash transfers and voting behavior. *Working paper*.

Nieto, C. (2014). *Migración haitiana a Brasil: Redes migratorias y espacio social transnacional*. CLACSO Buenos Aires.

Ohme, J. (2021). Algorithmic social media use and its relationship to attitude reinforcement and issue-specific political participation–the case of the 2015 european immigration movements. *Journal of Information Technology & Politics*, 18(1):36–54.

Ortega, F. and Polavieja, J. G. (2012). Labor-market exposure as a determinant of attitudes toward immigration. *Labour Economics*, 19(3):298–311.

Pacheco, J. S. (2008). Political socialization in context: The effect of political competition on youth voter turnout. *Political Behavior*, 30:415–436.

Pacifico, A. P., Ramos, E. P., Abreu Bastista Claro, C. d., and Cavalcante de Farias, N. B. (2015). The migration of haitians within latin america: significance for brazilian law and policy on asylum and migration.

Parsons, C. and Vézina, P.-L. (2018). Migrant networks and trade: The vietnamese boat people as a natural experiment. *The Economic Journal*, 128(612):F210–F234.

Parsons, C. R. and Winters, L. A. (2014). International migration, trade and aid: A survey. In *International handbook on migration and economic development*, pages 65–112. Edward Elgar Publishing.

Paternoster, R., Brame, R., Mazerolle, P., and Piquero, A. (1998). Using the correct statistical test for the equality of regression coefficients. *Criminology*, 36(4):859–866.

Pellegrina, H. S. and Sotelo, S. (2021). Migration, specialization, and trade: Evidence from brazil's march to the west. Technical report, National Bureau of Economic Research.

Peri, G. and Requena-Silvente, F. (2010). The trade creation effect of immigrants: evidence from the remarkable case of spain. *Canadian Journal of Economics/Revue canadienne d'économique*, 43(4):1433–1459.

Peri, G. and Sparber, C. (2009). Task specialization, immigration, and wages. *American Economic Journal: Applied Economics*, 1(3):135–69.

Petherick, A., Goldszmidt, R., Kira, B., and Barberia, L. (2020a). Do brazil's COVID-19 government response measures meet the who's criteria for policy easing?

Petherick, A., Kira, B., Barberia, L., Furst, R., Goldszmidt, R., Luciano, M., and Majumdar, S. (2020b). Brazil's fight against COVID-19: risk, policies, and behaviours. *BSG Working Paper Series*, 36:1–49.

Petroni, F. and Serva, M. (2010). Measures of lexical distance between languages. *Physica A: Statistical Mechanics and its Applications*, 389(11):2280–2283.

Piyapromdee, S. and Spittal, P. (2020). The income and consumption effects of COVID-19 and the role of public policy. *Fiscal studies*, 41(4):805–827.

Pop-Eleches, G. (2008). From economic crisis to reform. In *From Economic Crisis to Reform*. Princeton University Press.

Power, T. J. (2009). Compulsory for whom? mandatory voting and electoral participation in brazil, 1986-2006. *Journal of Politics in Latin America*, 1(1):97–122.

Prates, I. and Barbosa, R. J. (2020). The impact of COVID-19 in brazil: labour market and social protection responses. *The Indian Journal of Labour Economics*, 63(1):31–35.

RAMIREZ, E. G. (2021). Mexico's parliament and other political institutions.

Rauch, J. E. (1999). Networks versus markets in international trade. *Journal of international Economics*, 48(1):7–35.

Rauch, J. E. and Trindade, V. (2002). Ethnic chinese networks in international trade. *Review of Economics and Statistics*, 84(1):116–130.

Rawlinson, T., Siqueira, A. M., Fontes, G., Beltrão, R. P. L., Monteiro, W. M., Martins, M., Silva-Júnior, E. F., Mourão, M. P. G., Albuquerque, B., Alecrim, M. d. G. C., et al. (2014). From haiti to the amazon: public health issues related to the recent immigration of haitians to brazil. *PLoS Neglected Tropical Diseases*, 8(5):2685.

refworld. Brazil and haiti: Situation of haitians in brazil, including rights and obligations; permanent resident status; documents issued to haitians, including foreigner identity cards (cédula de identidade de estrangeiro, cie); treatment of haitians in brazil, including access to employment and education, state protection and support services (2010-june 2018). *refworld*.

Romero, S. (2016). Conservative's star rises in brazil as polarizing views tap into discontent. *The New York Times*.

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

Ruiz, I. and Vargas-Silva, C. (2018). Differences in labour market outcomes between natives, refugees and other migrants in the uk. *Journal of Economic Geography*, 18(4):855–885.

Ruseishvili, S., de Carvalho, R. C., and Nogueira, M. F. S. (2018). Construção social do estado de emergênia e a governança das migrações o decreto estadual no 24.469-e como divisor de aguas. *Migrações Venezuelanas. Universidade Estadual de Campinas*.

Russell, L. and Sun, C. (2020). The effect of mandatory child care center closures on women's labor market outcomes during the COVID-19 pandemic.

Saarimaa, T. and Tukiainen, J. (2016). Local representation and strategic voting: Evidence from electoral boundary reforms. *European Journal of Political Economy*, 41:31–45.

Sachs, J. D., Karim, S. A., Aknin, L., Allen, J., Brosbøl, K., Barron, G. C., Daszak, P., Espinosa, M. F., Gaspar, V., Gaviria, A., et al. (2020). Lancet COVID-19 commission statement on the occasion of the 75th session of the un general assembly. *The Lancet*, 396(10257):1102–1124.

Sanders, M. S. (2001). Uncertainty and turnout. *Political Analysis*, 9(1):45–57.

Schattschneider, E. E. (1975). *The semisovereign people: A realist's view of democracy in America*. Wadsworth Publishing Company.

Serva, M. and Petroni, F. (2008). Indo-european languages tree by levenshtein distance. *EPL (Europhysics Letters)*, 81(6):68005.

Sewordor, E., Esnard, A.-M., Sapat, A., and Schwartz, L. (2019). Challenges to mobilising resources for disaster recovery and reconstruction: perspectives of the haitian diaspora. *Disasters*, 43(2):336–354.

Silva, C. X. d. (2019). Imigração haitiana: um estudo sobre o estabelecer do imigrante na cidade no contexto histórico e social de globalização.

Silva, J. S. and Tenreyro, S. (2006). The log of gravity. *The Review of Economics and statistics*, 88(4):641–658.

Small, K. A. and Hsiao, C. (1985). Multinomial logit specification tests. *International economic review*, pages 619–627.

Solt, F. (2010). Does economic inequality depress electoral participation? testing the schattschneider hypothesis. *Political behavior*, 32(2):285–301.

Soss, J. and Jacobs, L. R. (2009). The place of inequality: Non-participation in the american polity. *Political Science Quarterly*, 124(1):95–125.

Sousa, R. (2015). Imigração haitiana no brasil. *Mundo Educacão*.

Steinmayr, A. (2021). Contact versus exposure: Refugee presence and voting for the far right. *Review of Economics and Statistics*, 103(2):310–327.

Stockemer, D. (2017). What affects voter turnout? a review article/meta-analysis of aggregate research. *Government and Opposition*, 52(4):698–722.

Strozzi, C. (2015). How migration fosters trade.

Tabellini, M. (2020). Gifts of the immigrants, woes of the natives: Lessons from the age of mass migration. *The Review of Economic Studies*, 87(1):454–486.

Tai, S. H. (2009). Market structure and the link between migration and trade. *Review of World Economics*, 145(2):225–249.

The Economist. The way forward on immigration to the west. *The Economist*.

The New York Times. Jair bolsonaro wins Brazil's presidency, in a shift to the far right. *The New York Times*.

The Organization for Economic Co-operation and Development (OECD) (2020). Coronavirus (COVID-19): Sme policy responses.

Tribunal Regional Eleitoral-SP. Voto obrigatório e voto facultativo saiba o que diz a constituição e quais as consequências de não votar. *Tribunal Regional Eleitoral*.

Uebel, R. R. G. (2017). International migrations to brazil in the 21st century: Profile, outlook and trends. *Population Review*, 56(1).

Uebel, R. R. G. (2018). Política externa migratória brasileira: das migrações de perspectiva à hiperdinamização das migrações durante os governos lula da silva e dilma rousseff.

Uebel, R. R. G. (2020). Brazilian foreign policy for immigrants and refugees: New concepts and ethical issues. *Revista Brasileira de Sociologia*, 8(19):80–97.

Van Dyck, B. (2014). Why party organization still matters: The workers' party in northeastern brazil. *Latin American Politics and Society*, 56(2):1–26.

Vasconcelos, I. and Santos, S. (2018). Refugiados em roraima? instituições, papéis e a competição pelas categorias. *Migrações Venezuelanas. Campinas, SP: Núcleo de Estudos de População "Elza Berquó"–Nepo/Unicamp*.

Verba, S., Schlozman, K. L., and Brady, H. E. (1995). *Voice and equality: Civic voluntarism in American politics*. Harvard University Press.

Verschuur, J., Koks, E. E., and Hall, J. W. (2021). Global economic impacts of COVID-19 lockdown measures stand out in high-frequency shipping data. *PloS one*, 16(4):e0248818.

Wagner, D., Head, K., and Ries, J. (2002). Immigration and the trade of provinces. *Scottish Journal of Political Economy*, 49(5):507–525.

Wang, X. and Badman, R. P. (2016). A multifaceted panel data gravity model analysis of peru's foreign trade. *arXiv preprint arXiv:1612.01155*.

Wejsa, S. and Lesser, J. (2018). Migration in brazil: the making of a multicultural society. *The Online Journal of Migration Policy Institute*.

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica: journal of the Econometric Society*, pages 817–838.

White, R. and Tadesse, B. (2008). Immigrants, cultural distance and us state-level exports of cultural products. *The North American Journal of Economics and Finance*, 19(3):331–348.

Wichmann, S., Holman, E. W., Bakker, D., and Brown, C. H. (2010). Evaluating linguistic distance measures. *Physica A: Statistical Mechanics and its Applications*, 389(17):3632–3639.

Wichmann, S., Holman, E. W., and Brown, C. H. The asjp database (version 19). Technical report.

Wiesehomeier, N. and Benoit, K. (2009). Presidents, parties, and policy competition. *The Journal of Politics*, 71(4):1435–1447.

Wiesehomeier, N., Singer, M., and Ruth-Lovell, S. (2021). Political representation, executives, and political parties survey: Data from expert surveys in 18 latin american countries, 2018-2019. prepps latam v2.

Wooldridge, J. M. (2002). Econometric analysis of cross section and panel data mit press. *Cambridge, MA*, 108(2):245–254.

Yates, C. Haitian migration through the americas: A decade in the making. *Migration Policy Institute*.

Zhang, X. N., Wang, W. W., Harris, R., and Leckie, G. (2020). Analysing interprovincial urban migration flows in china: A new multilevel gravity model approach. *Migration Studies*, 8(1):19–42.

Zuckerman, A. S. (2005). *The social logic of politics: Personal networks as contexts for political behavior*. Temple University Press.