From the holdings of the Department of Special Collections , Princeton University Library, Princeton, NJ

These documents are either in the public domain, or can only be used for educational and research purposes (Fair use) as per U.S. Copyright law (text below). By accessing this file, all users agree that their use falls in the public domain or within fair use as defined by the copyrightlaw. We do not charge any permission or usage fees for the publication of images of material inour collections, including those provided by Princeton University Library, via our website, catalog, or directly from staff.For more information regarding copyright, credit, and citation please review the information onthe Copyright, Credit and Citation Guidelines page of our website.
https://library.princeton.edu/special-collections/policies/copyright-credit-and-citation-guidelines. If you have any questions, including how to order higher-quality images from the collection, contact Special Collections Public Services via our Ask Us! Form. https://library.princeton.edu/special-collections/ask-us

---

Doctoral Dissertations, 2011-2026 Economics

2020

Essays in Migration Economics

Porcher, Charly

Princeton, NJ : Princeton University

---

http://arks.princeton.edu/ark:/88435/dsp01pc289n140

U.S. Copyright law: The copyright law of the United States (Title 17, §108, United States Code) governs the making of photocopies or other reproductions of copyrighted material. Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction. One of these specified conditions is that the photocopy or reproduction is not to be used for any purpose other than private study, scholarship, or research. If a user makes a request for, or later uses, a photocopy or reproduction for purposes in excess of fair use that user may be liable for copyright infringement. This institution reserves the right to refuse to accept a copying order if, in its judgment, fulfillment of the order would involve a violation of copyright law. https://www.copyright.gov/title17/title17.pdf

Essays in Migration Economics

Charly Porcher

A Dissertation

Presented to the Faculty

of Princeton University

in Candidacy for the Degree

of Doctor of Philosophy

Recommended for Acceptance

by the Department of

Economics

Adviser: Professor Esteban Rossi-Hansberg

September 2020

© Copyright by Charly Porcher, 2020.

All rights reserved.

Abstract

This dissertation studies the economic determinants of migration decisions. Chapter 1, joint with Eduardo Morales and Thomas Fujiwara, examines whether workers' limited migration responses to local shocks are due to lack of information about the potential gains from regional migration. We analyze the mobility decisions of all formally employed workers in Brazil over 15 years. We derive model-based moment conditions to test for the content of migrants' information sets. First, we find that individuals located in regions with a better access to internet and a higher population density have richer information sets overall. Second, information is concentrated on few neighboring regions and on regions with population sharing similar education and racial composition. Third, individuals identifying as white and with a higher educational attainment feature richer information sets.

Chapter 2 explores the consequences of this imperfect and heterogeneous information structure for the spatial allocation of economic activity and welfare. I develop a quantitative dynamic model of migration with costly information acquisition and local information sharing. Agents optimally acquire more information about nearby locations and learn about other locations from the migrants around them. After estimating the model from the Brazilian administrative data, I evaluate the counterfactual effect of the roll-out of broadband internet in Brazil. By allowing workers to make better mobility choices, expanding internet access increases average welfare, reduces migration flows by limiting mistakes and reduces the cross-sectional dispersion in earnings.

Chapter 3, joint with Hannah Rubinton and Clara Santamaría, studies the origins of large cities' attractiveness for migrants' career prospects. We investigate whether the different composition of establishment sizes between cities can help explain the city-size earnings premium. Using administrative data from Spain, we find that nearly a third of the short-term gains of moving to a city twice larger can be explained by a

transition to a better-paying larger establishment. In contrast, only a small fraction of the medium-term gains of accumulating experience in a large city can be attributed to a faster transition to larger establishments. We conclude that the establishment size composition is fundamental for understanding the short-term gains of moving to a larger city.

Acknowledgements

I want to express my deepest gratitude to my advisors Esteban Rossi-Hansberg and Eduardo Morales for their continuous inspiration, support and guidance. They have helped me through countless iterations of my research, and their encouragement and insights have been invaluable during my journey to become a researcher. It has been my pleasure and fortune to learn the craft from them.

My development as an economist has also benefited greatly from the advice of many other members of the Princeton faculty and staff. I am obliged to Thomas Fujiwara, Gene Grossman, Adam Kapor, Jakub Kastl, Adrien Matray, Steve Redding, Richard Rogerson, Chris Sims, all of whom have influenced my research during my time at Princeton. This work also benefited greatly from the stimulating environment provided by the International Trade Workshop at Princeton University. I thank all their participants. I also thank Laura Hedden and Sharon Ernst for always ensuring that things run smoothly. I gratefully acknowledge financial support from the International Economics Section and the Industrial Relations Section at Princeton University.

I thank my classmates and officemates for their companionship and support throughout the past six years. I would especially like to thank Federico Huneeus, Korhan Koçak, Mathilde Lemoigne, Hugo Lhuillier, Anders Nielsen, Hannah Rubinton and Clara Santamaría for their constant encouragement and unbeatable friendship. An extra special thanks is reserved to my great friends Adrien Bilal, Damien Capelle and Maurício Matsumoto for always reminding me that life as a researcher can and should be fun.

Last but not least, my sincere gratitude to my family for their unwavering support and affection: Annie Bruneau, Jean-Luc Porcher and Benjamin Porcher.

To my parents and my brother.

Contents

Abstract iii

Acknowledgements v

List of Tables xi

List of Figures xiii

1 A Revealed-Preference Approach to Measuring Information Frictions in Migration Decisions 1

1.1 Introduction 1

1.2 Data and Suggestive Evidence 8

1.3 Empirical Framework 21

1.4 Estimation of Migration Elasticity 35

1.5 Identification of Information Frictions 36

1.6 Results 38

1.7 Concluding Remarks 54

2 Migration with Costly Information 57

2.1 Introduction 57

2.2 A Model of Migration with Costly Information Acquisition and Local Information Sharing 66

2.3 Data Sources 86

2.4 Estimating the Model 88

2.5 Testing Predictions of the Model 100
2.6 Counterfactual Exercises 111
2.7 Conclusion 118

3 The Role of Establishment Size in the City-Size Earnings Premium in Spain 120

3.1 Introduction 120
3.2 Data Description 125
3.3 Establishment-Size Composition and the City-Size Earnings Premium 127
3.4 The Role of Establishment Size in the City-Size Earnings Premium . 134
3.5 Conclusion 142

A Appendix for "A Revealed-Preference Approach to Measuring Information Frictions in Migration Decisions" 143

A.1 Proofs and Additional Derivations 144
A.1.1 Section 1.4: Additional Derivations for the Myopic Case . . . 144
A.1.2 Section 1.4: Additional Derivations for the Forward Looking Case 145
A.1.3 Individual-Specific Information Sets - Partial Identification with Moment Inequalities 150
A.2 Descriptive Statistics and Additional Suggestive Evidence on Information Frictions 166
A.2.1 Descriptive Statistics 166
A.2.2 Additional Suggestive Evidence on Information Frictions . . . 179
A.3 Additional Results on Testing for the Content of Information Sets . . 187
A.3.1 Derivation of the Overidentification Test 187
A.3.2 Implementation of the Overidentification Tests 189
A.3.3 Additional Tests of the Content of Workers' Information Sets . 195

B Appendix for "Migration with Costly Information" 197

B.1 Appendix to Section 2.2 . . . 198

B.1.1 Proofs . . . 198

B.1.2 Alternative Assumptions on the Invertibility of Wages and Population . . . 212

B.1.3 Solution Algorithm . . . 220

B.1.4 Accuracy of the Solution Algorithm . . . 224

B.2 Appendix to Section 2.3 . . . 228

B.2.1 Construction of the Main Sample . . . 228

B.2.2 Additional Descriptive Statistics on Migration Patterns . . . 229

B.3 Appendix to Section 2.4 . . . 233

B.3.1 Algorithm for Simulated Method of Moments . . . 233

B.3.2 Construction of the Instrument for Internet Access . . . 235

B.4 Appendix to Section 2.5 . . . 238

B.4.1 Fit of Migration Flows . . . 238

B.4.2 Heterogeneous Migration Elasticities in the Model with No Information Frictions . . . 238

B.4.3 List of Local Shocks . . . 238

B.4.4 Replicating Local Shocks in the Model . . . 240

B.5 Additional Figures . . . 242

B.5.1 Cardell Distribution . . . 242

C Appendix for "The Role of Establishment Size in the City-Size Premium in Spain" 243

C.1 Data Appendix . . . 244

C.1.1 Data Definitions . . . 244

C.1.2 Summary statistics . . . 246

C.1.3 Firm and Establishment Co-workers . . . 248

C.2 Additional Results 251
C.2.1 A Closer Look into Mover Regressions 251
C.2.2 Heterogeneity in the Co-worker Earnings Premium Across Cities 253

Bibliography **257**

List of Tables

1.1 Description of Positive Local Labor Demand Shocks 14

1.2 Estimates of the Elasticity of Migration With Respect to Wages by OLS and with Different Instrumental Variables 41

1.3 Testing Content of Information Set - All Groups 43

1.4 Testing Content of Information Set - Heterogeneity Across Destinations 44

1.5 Testing Content of Information Set - Heterogeneity Across Groups . . 49

1.6 Testing Content of Information Set - Interactions 53

2.1 Estimated Parameters 93

2.2 Determinants of Information Costs 95

2.3 Determinants of Migration Costs 100

2.4 Heterogeneity of Migration Elasticities in the Data and in the Model 104

3.1 City-Size Earnings Premium, Co-Worker Earnings Premium and City-Size coworker premium 129

3.2 City Size Co-worker Premium by Sector 134

3.3 The Role of Transitions to Larger Establishments in the Average City-Size Earnings Premium 136

3.4 The Role of Transitions to Larger Establishments for Short-Term and Medium Term City-Size Earnings Premium 140

A.2.1 Average Fixed Effect Estimates by Year 166

A.2.2Average Fixed Effect Estimates by Demographic Group 167
A.2.3Average Fixed Effect Estimates by Sector 167
A.2.4Average Fixed Effect Estimates by Microregion 169
A.2.5RAIS: Top / Bottom 10 Microregions by Population Size 172
A.2.6Census: Top / Bottom 10 Microregions by Population Size 173
A.2.7Distribution of employees by Sector (mean over years) 173
A.2.8Mean Population and Income over Years, Transitions 176
A.2.9Top/ Bottom 10 most attractive origin microregions 177
A.2.10Top/ Bottom 10 most attractive sectors 178
A.3.1Testing Content of Information Set - Static Model 196

B.2.1Mean and Quantiles of Main Variables 232
B.4.1Description of Positive Local Labor Demand Shocks 240

C.1.1Summary Statistics 247
C.1.2City Size Co-worker Premium: Establishments vs Firms 249
C.1.3Co-worker Earnings Premium: Establishments vs Firms 250
C.2.1Selection of Movers and Worker Fixed Effects 252
C.2.2Non-random Moves across Firm Size Categories 253
C.2.3Co-workers Earnings Premium and City Size 256

List of Figures

1.1 Population in 2014 by Microregion 12
1.2 Heterogeneity in Migration Responses: Labor Market Shock in Altamira 18
1.3 Heterogeneity in Migration Responses: Pooled Labor Market Shocks . 20

2.1 Sequence of actions during any period t 70
2.3 Estimated Migration Costs with and without Information Cost 98
2.4 Delay in the Migration Response to the Local Shock in Ipojuca . . . 109
2.5 Welfare Gains by Mesoregion after Setting All Information Costs to Zero 112
2.6 Welfare Gains by Mesoregion after Expansion of Internet Access . . . 116

3.1 Co-worker Earnings Premium by City 132

A.2.1 Heterogeneity in Migration Responses: Labor Market Shock in Suape 179
A.2.2 Heterogeneity in Migration Responses: Labor Market Shock in Porto Velho 181
A.2.3 Heterogeneity in Migration Responses: Labor Market Shock in Medio Mearim 183
A.2.4 Heterogeneity in Migration Responses: Labor Market Shock in Quirinópolis 185

B.1.1 Sequence of actions during any period t 216
B.1.2 Simulation Error 225

B.2.1Migration rates across mesoregions by Year over 2000-2014 229
B.2.2Migration rates across mesoregions by Year, by age, over 2000-2014 . 230
B.2.3Migration rates across mesoregions by Year, by gender, over 2000-2014 230
B.2.4Migration rates across mesoregions by Year, by race, over 2000-2014 . 231
B.2.5Migration rates across mesoregions by Year, by educational attainment, over 2000-2014 231
B.2.6Population Density by Mesoregion - Average over 2000-2014 232
B.3.1Change in internet access from 2000-2007 to 2008-2014 and new backbone cables 237
B.4.2Average Delay in the Migration Response to the 10 Local Shocks . . . 240
B.5.1Distribution Cardell with dispersion $\beta = 0.5$, vs. Gumbel. Source: Dasgupta and Mondria (2014). 242

Chapter 1

A Revealed-Preference Approach to Measuring Information Frictions in Migration Decisions

joint with Thomas Fujiwara and Eduardo Morales

1.1 Introduction

Labor demand shocks differ widely across regions within countries (Bartik, 1991; Blanchard and Katz, 1992).[1] By migrating, workers thus have the capacity to avoid negative employment and wage shocks in their region of residence and to benefit

[1]This cross-regional heterogeneity in labor demand shocks may be due to regions' differential exposure to international trade competition (Topalova, 2010; Kovak, 2013; Autor et al., 2013; Adão, 2016; Dix-Carneiro and Kovak, 2017; Pierce and Schott, 2018; Dix-Carneiro and Kovak, 2019) and changes in technology (Beaudry et al., 2010; Autor and Dorn, 2013; Bustos et al., 2016; Acemoglu and Restrepo, 2017). Cross-regional differences in housing net worth (Mian and Sufi, 2014), government spending (Nakamura and Steinsson, 2014), business cycles (Beraja et al., 2019), the location of firms (Greenstone et al., 2010), and natural resource discovery (Feyrer et al., 2017) have also been documented as sources of regional dispersion in labor demand shocks.

from positive labor demand shocks in other regions (Bryan et al., 2014). However, in many occasions, workers' migration patterns do not respond to these regional labor demand shocks. A common explanation for this lack of response is the existence of very high migration costs (Kennan and Walker, 2011), which have been shown to have important implications for aggregate labor productivity and welfare.[2]

In this paper, we study the extent to which workers' limited migration responses to local labor demand shocks are due to their lack of information about their potential net gains from regional migration. Using data for the period 2000 to 2014 on the location of work of every legally employed Brazilian worker, we document how workers' individual characteristics and the characteristics of the place in which they reside affect the extent to which they are informed about the annual wages and amenity levels in every Brazilian local labor market. Our results show that gravity in terms of geographic, educational and racial distance, as well as the degree of internet penetration in the destination and in the region of residence of a worker crucially determine the extent to which such worker is informed about the net gains of migrating to any other Brazilian region. We also find that workers with a higher educational attainment and who identify as white have richer information sets. Conversely, neither the total population, the population density in the destination, nor the average income level of a location seem to determine the information to which their residents have access to.

Our analysis combines both reduced-form and a revealed-preference structural approach to the identification of information frictions in migration.

In our reduced-form evidence, we measure the *delay* with which certain groups of workers react to localized positive labor demand shocks. To perform this analysis, we identify a set of Brazilian localities that experienced a labor demand shock during our

[2]For evidence on the lack of labor mobility in reaction to regional labor demand shocks, see Adão et al. (2019) and several of the papers cited in footnote 9. For quantifications of the aggregate implications of workers' lack of labor mobility, see Morten and Oliveira (2018); Bryan and Morten (2018); Caliendo, Opromolla, Parro and Sforza (2018); Caliendo, Dvorkin and Parro (2019).

sample period. As an example, one of these localities is Altamira, located in the state of Pará. In 2010, the "Consorcio Constructor Belo Monte" started the construction of the Belo Monte dam. The population of Altamira rose from below 13,000 workers in 2009 to above 60,000 workers in 2014. However, not all the immigrants that moved into Altamira did so at the same time: *ceteris paribus*, workers initially located in regions geographically close to Altamira and in regions with a relatively larger share of households with internet access moved earlier. Conversely, we see no differences in the speed with which migrants reacted to the shock depending on their age, gender, and education level. We observe similar patterns when studying the delay with which workers moved to other regions for which we could document positive labor demand shocks similar to that of Altamira in 2010.[3]

The reduced-form exercises described above are suggestive of the presence of information frictions affecting workers' migration decisions and that are heterogeneous across workers depending on the characteristics of the place in which they live. However, given the reduced-form nature of those analyses, the assumptions under which one can use these reduced-form facts to infer the nature and strength of information frictions in migration are unclear. To address this shortcoming, we present in the second part of the paper a structural analysis of information frictions.

Our structural analysis is based on a dynamic discrete choice model of labor mobility across regions and sectors. As in Artuç and McLaren (2015) and Caliendo et al. (2019), workers in our model decide in every period both the sector and the region of employment. When doing so, workers trade off higher real wages and amenities for potential migration and sector-switching costs, and take into account the impact of their current decisions on their subsequent choices and the corresponding stream of

[3]Specifically, other region-year pairs experiencing a sudden positive labor demand shock during our sample period are: Porto Velho (in the state of Rondônia) in 2002, when the construction of the Santo Antonio dam was initiated; Ipojuca (in the state of Pernambuco) in 2007, when a $14 billion refinery, a $2 billion petrochemical plant and a $8 billion shipyard started their construction following the discovery of massive offshore oil beds off the coast of Ipojuca.

possible wages and amenity flows. Importantly, when deciding on their optimal sector and region in which to be employed, workers may have limited information about the wages and amenity levels that they would enjoy upon choosing any of the sector-region pairs in their choice set. Besides being potentially heterogeneous in their information sets, workers may also have region and sector-specific idiosyncratic preferences that are unobserved to the researcher. We assume that workers' unobserved idiosyncratic preferences for any given region-sector pair are such that the probability that any given worker is employed in a particular region and sector in any given year follows a nested logit model.

While the data we have access to contain worker-specific information on the realized wages for those sector-region pairs in which the corresponding worker is effectively employed in any given year, they naturally do not contain information on the hypothetical wages that a worker could have obtained if she had chosen a region or sector different than that observed in the data. However, estimating our dynamic discrete choice model and using it as basis to infer the information set that different types of workers have at each point in time requires constructing a proxy for those unobserved wage levels. When doing so, we allow the wages that any worker would obtain if employed in a particular sector and region to depend on individual worker characteristics (i.e. age, gender, education level, and region- and sector-specific experience) with year-specific coefficients, on region- and year-specific fixed effects, on worker-specific fixed effects, and on an unexpected region-, sector-, year- and individual-specific unobserved term.[4]

[4]This wage equation does not include a region-, sector- and individual-specific term that is constant over time, unobserved to the researcher, and potentially known by workers when deciding on their preferred region and sector of employment. See Kennan and Walker (2011); Dix-Carneiro (2014); Traiberman (2019) for specifications of wage equations that allow for such a term in the context of dynamic discrete choice models of labor mobility. Allowing for such a component of wages poses estimation challenges when, as it is our case, one imposes only a minimum content requirement on agents' information sets.

The purpose of our structural model is not to predict the general equilibrium impact of counterfactual labor demand shocks and informational interventions, but rather to characterize the extent to which workers are uncertain about labor market conditions in different regional labor markets in Brazil. Specifically, given our model and our data on Brazilian workers' location choices, we show how data on migration flows can be used to infer the information that workers have about the labor market conditions and amenity levels that these workers would enjoy if they were to move to any regional labor market in Brazil. Methodologically, our approach is thus similar to that in Dickstein and Morales (2018) in that we use model selection tests for moment conditions to evaluate the null hypothesis that certain covariates observed in our data belonged to workers' information sets at the time at which they had to take the migration decisions that we study and aim to understand. However, in our main specification, we rely on moment equalities, while Dickstein and Morales (2018) used inequality models.[5]

Our paper is related to several strands of the literature. First, it relates to papers that aim to understand the observed lack of labor mobility in the face of large variations in income levels and amenities across locations. A strand of the migration literature introduces static models of labor location that build formally on Roy (1951) and that rationalize workers' lack of perfect mobility on the basis of them having location-specific idiosyncratic preferences.[6] Building on this literature, a related

[5]In Appendix A.1.3, we derive moment inequality procedures that extend those in Dickstein and Morales (2018) in two dimensions. First, we show how to deal with discrete choice settings in which agents' choice sets include more than two choices; specifically, in our application, workers may choose their preferred option from a choice set that may include more than 500 choices. Second, we show how to transform the odds-based moment inequalities introduced in Dickstein and Morales (2018) in a way that the resulting moment inequality is linear in the arguments entering the agents' static utility function. This allows us to account for a broad set of fixed effects in workers' utility. In our application, accounting for such set of fixed effects is important, as it allows us to infer workers' information sets while being fairly agnostic about the factors determining the migration costs that workers face when deciding where to migrate. The empirical implementation of these moment inequalities is left for future work.

[6]For early examples of this approach in the international migration literature, see, for example, Borjas (1987, 1999); Grogger and Hanson (2011). For applications to local labor markets, see, for

strand of the literature introduces static frameworks that incorporate additional barriers to migration in the form of fixed migration costs.[7] Starting with Kennan and Walker (2011), a more recent literature has introduced both partial and general dynamic equilibrium analysis of migration featuring both one-time migration costs and idiosyncratic worker preferences for locations.[8] We build a dynamic partial equilibrium model of migration decisions and depart from the previous literature in that we do not fully specify the content of workers' information sets at the time at which migration decisions are taken. Instead, we impose only a minimal content requirement on migrants' information sets and use revealed-preference arguments to infer the information frictions that workers face when deciding where to move.

Our paper also contributes to the literature studying the relevance of information frictions for migration. Prominent studies have analyzed how experimental variation in access to information about payoff-relevant variables affects migration decisions.[9] Some examples include Bryan et al. (2014), Jensen and Miller (2017), Bergman et al. (2020), Bergman et al. (2019), Baseler (2019).[10] More recently, Kaplan and Schulhofer-Wohl (2017) discussed the structural identification of information frictions when migration decisions are potentially subject to other frictions. Porcher (2020) develops a dynamic general equilibrium model of migration in which infor-

example, Dahl (2002); Moretti (2011); Busso et al. (2013); Diamond (2016a); Redding (2016); Suárez Serrato and Zidar (2016); Fajgelbaum et al. (2019).

[7]See, for example, Morten and Oliveira (2018); Monte et al. (2018); Tombe and Zhu (2019); Fan (2019). For an early application of this approach to within-city mobility, see Ahlfeldt et al. (2015).

[8]For partial equilibrium analysis, see, for example, Bishop and Murphy (2011); Bayer et al. (2016). For general equilibrium analysis, see Artuç and McLaren (2015); Caliendo et al. (2019). Instead of modeling migration costs as a one-time cost, Desmet et al. (2018) model these as a permanent flow utility cost.

[9]More broadly, our paper also connects to the literature studying the relevance of information frictions for other economic decisions. Examples using experimental evidence include Jensen (2010), Guiteras et al. (2015), Kendall et al. (2015), Miller and Mobarak (2015), Aker et al. (2017), and BenYishay and Mobarak (2018). Other contributions exploiting quasi-experimental evidence are Jensen (2007), Aker (2010), Goyal (2010), Brown and Goolsbee (2002), Koudijs (2015, 2016), Hastings et al. (2015), Steinwender (2018), Allende et al. (2019) . Allen (2014) and Berman et al. (2019) study the structural identification of information frictions in international trade.

[10]See also Baseler (2016), Wilson (2018) who provide suggestive evidence of the importance of information frictions in migration choices and Shrestha (2017) for a survey of migrants' expectations.

mation is costly. In his model, agents are rationally inattentive and make decisions with incomplete information. Applying his model to the Brazilian economy, he shows how information frictions can be separately identified from, in particular, migration costs. Our contribution relative to this literature is to propose a revealed-preference method to infer the importance of information frictions without the need to observe exogenous determinants of agents' information sets. As such, our revealed- preference approach is applicable to many settings and can allow for rich heterogeneity in agents' information sets.

Finally, our paper also contributes to the literature studying the use of choice data to identify agents' preferences when their expectations of the choice characteristics are unobserved. Alternatives suggested in this literature to deal with unobserved agents' expectations include assuming that the researcher can construct a perfect proxy for agents' expectations, as in Manski (1991, 1993, 2007), Ahn and Manski (1993) and Cunha and Heckman (2007). A second approach is to directly observe a measure of agents' subjective expectations, as in Manski (2004), De Bruin et al. (2011), Delavande et al. (2011), van der Klauuw (2012) and Armantier et al. (2013). A third approach is to assume that agents' expectations are rational and that the researcher observes a subset of agents' information sets. Examples include Scott (2013), Dickstein and Morales (2013) Dickstein and Morales (2018), Morales et al. (2017), Traiberman (2019), Humlum (2020) and Diamond et al. (2019). We follow the third approach and show how it can be extended to test which variables belong to agents' information sets.

The paper is organized as follows. In Section 2.3, we describe the data and provide suggestive evidence of the role of information frictions in migration decisions in Brazil. In Section 2.2, we build the empirical framework. Section 1.4 discusses the estimation strategy for the elasticity of migration with respect to wages. Section 1.5 develops the

identification of information frictions and the testing method. Section 1.6 presents the results.

1.2 Data and Suggestive Evidence

In this section, we describe our sources of data, provide descriptive statistics on the migration patterns we observe, and present evidence suggestive of the relevance of information frictions in determining workers' migration decisions. We also discuss alternative mechanisms that may generate similar migration behavior. This discussion informs the specification of the model introduced in Section 2.2, which we use as basis for an approach aimed at controlling for these alternative mechanisms when measuring the importance of information frictions for workers' migration decisions.

1.2.1 Data Sources and Descriptive Statistics

We combine multiple sources of data. Our main data source is the *Relação Anual de Informações Sociais* (RAIS), which is collected by the Brazilian Ministry of Labor and contains matched employer-employee information for every formally employed worker in Brazil. We use the data corresponding to the period 1993 to 2014. For every formal job and year, we exploit information on the day of accession into the job and the day of separation (if either of them took place during the corresponding year), the average monthly wage, the number of hours stipulated in the contract, the 2-digit occupation (according to the *Classificação Brasileira de Ocupações*, CBO), and certain characteristics of both the plant at which the worker is employed and of the worker herself. Specifically, we use information on the micro-region in which the plant is located and its main 2-digit industry of production (according to the

Classificação Nacional de Atividades Econômicas, CNAE), as well as information on the workers' gender, age, race and level of education.[11]

Our data contains no information on the location of residence of a worker before their first job in the formal sector. Consequently, in our analysis, we focus on the migration decisions of workers and do not model the decision of potential workers of whether to enter the labor force or acquire college education. For this reason, we limit our data to observations that correspond to workers that are over 25 years of age, since, for the majority of the population, educational decisions are taken before this age. Similarly, we do not model the retirement decision of workers and, consequently, we limit our data to observations that correspond to workers that are below 65 years of age.

Because RAIS only contains information on the formal employment of workers in Brazil, we have no information on the location of workers that do not hold a formal job in a given year. These workers may be employed in the informal sector, self-employed, unemployed, or out of the labor force. In our model, we account for the possibility that the precise employment status of a worker as well as her location are unobserved. However, given that our results will naturally capture only the incidence of informational frictions for migration decisions of workers employed in the formal sector, we limit our data to workers that have a sufficiently close labor relationship with the formal sector; specifically, we limit our sample to workers appearing in our sample for at least 7 years between 2000 and 2014.

It is not infrequent that workers in the sample will appear as performing multiple different jobs in the same year. In order to obtain a dataset in which each unit of observation corresponds to a worker and a year, we assign to each worker-year specific pair the location, sector and occupation corresponding to the job that the worker held

[11] Brazilian microregions are groups of municipalities that span the entirety of the Brazilian territory. During our sample period, there were 557 microregions which may themselves grouped into 136 mesoregions.

for the longest period of time during the corresponding year. However, to determine the total labor income of a worker in a year, we add the labor income earned by the worker in every job in which, according to our data, this worker has been employed in the corresponding year.[12]

Although the data to which we have access contains information on all variables going back 1993, a discontinuity in the way the wage information is reported prior to the year 1999 implies that we will only study the migration decision of the workers in our sample during the period 2000 to 2014. We however use the information for the years 1993 to 1999 to measure the experience that each worker had in each sector, occupation and microregion in the year 2000. For those workers whose first year in the sample is posterior to the year 1999, we assume that their experience in every sector, occupation and microregion in their first year appearing in our data is equal zero. Given these initial experience levels for every worker, we use the information on their sector, occupation and location of formal employment in every subsequent year to update their experience levels.

Besides the information on workers' labor market histories contained in the RAIS database, we also use information on the population of each municipality in each year between 2000 and 2014 from the population census collected by *Instituto Brasileiro de Geografia e Estatistica* (IBGE). We compute a measure of population by microregion by aggregating the population of all municipalities included in the corresponding microregion.

Finally, we collect information on the degree of internet penetration by microregion. Specifically, from the *Agência Nacional de Telecomunicações*, ANATEL, a government agency in charge of regulating and supervising telecommunications in Brazil,

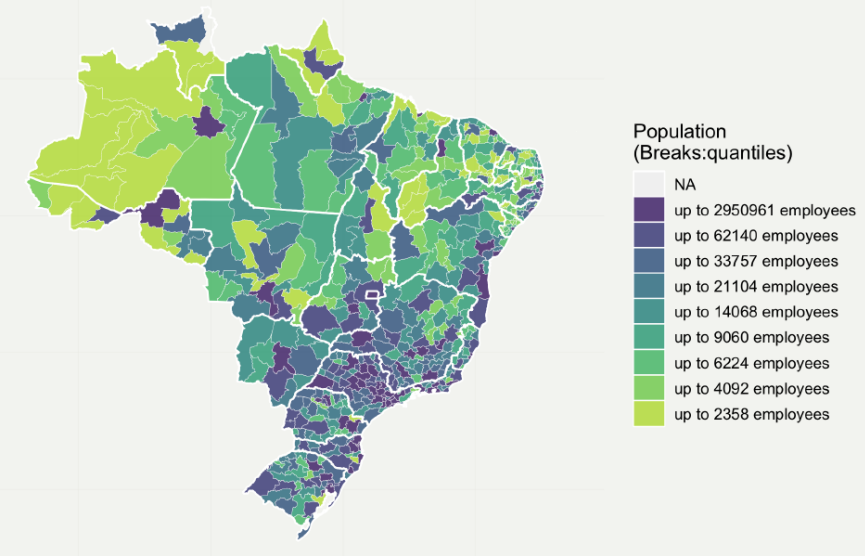
[12]To compute the total labor income of a worker associated to each job this worker has held, we transform the average monthly earnings reported for each job into a measure of average daily wages, and multiply this one by the total number of days between the day of accession and the day of separation into the job reported in the data. If no information on the day of accession or separation is reported, we assume that these ones are January 1 and December 31, respectively.

we obtain information on the number of broadband connections by municipality and year between 2007 and 2014. We use the population data obtained from the population census to construct a municipality-specific measure of the number of broadband connections per capita, and we use again the municipality-level population data to construct an equivalent variable at the microregion level as a population-weighted average of the per capita number of broadband connections in the municipalities included in the corresponding microregion. As we do not have access to information on the number of broadband internet connections for the years 2000 to 2006 and the available data indicates that the overall number of broadband connections per inhabitant is less than 1% in 2007, we assume that the number of broadband connections per capita equals zero in every microregion and year prior to 2007.

The resulting dataset includes 372,454,979 worker-year pairs that correspond to 45,958,805 workers. According to the RAIS data, the 557 microregions in which the geography of Brazil is divided had in 2014 on average close to 59,000 legal workers (the median microregion had close to 15,000 legal workers). The largest microregions by number of workers in our dataset were Sao Paulo (close to 3 million workers), Rio de Janeiro (approximately 1.7 million) and Belo Horizonte (slightly below 1 millions workers); these three regions accounted for 12.4%, 7.2% and 4.1% of the overall number of workers in the data in 2014, respectively. Figure B.2.6 shows the 2014 population by microregion; this map illustrates clearly that the most populated microregions are located in the South and along the coast. Conversely, the microregions located in the North and in the West are much less populated.

As indicated above, we classify each worker into a 2-digit sector and a 2-digit occupation. The total number of sectors in our data is thus 59, and the total number of occupations is 28. The sectors that, on average between 2000 and 2014, have the highest number of workers in the formal economy are: public administration (15.5%), retail trade (11.8%), service activities -such as legal, accounting and security-mainly

Figure 1.1: Population in 2014 by Microregion



Notes: Number of formal employees in each microregion in 2014. Darker colors indicate microregions with larger populations. The territories marked with white lines correspond to states. Source: RAIS

provided to businesses (8.81%). Similarly, the occupations with the largest number of workers in the database are: general and keyboard clerks (15.5%), personal service workers (9.76%), and cleaners and helpers (7.01%).

1.2.2 Suggestive Evidence: Delay in Migration Decisions

In this section, we provide suggestive evidence of the empirical relevance of the existence of information frictions affecting workers' migration decisions. We also identify certain factors that are potential determinants of the information that workers have about local labor demand shocks that take place in Brazilian labor markets other than the location of residence of each workers.

If the labor market of a location is perfectly competitive and aggregate labor supply in such market is not perfectly elastic, a positive labor demand shock in this location will translate into an increase in wages, which will attract workers from other locations. In a model with perfect information (e.g. Caliendo et al., 2019), the delay

with which workers would react to a positive labor demand shock in a location other than their location of residence is a function exclusively of the prior migration shares between the location of residence of the worker and that experiencing the positive labor demand shocks. We illustrate here that this property of models of migration decisions with perfect information does not hold in the data.

In order to empirically identify a microregion as having experienced a positive labor market demand shock, we require this microregion to verify three criteria. First, there must exist a year such that the minimum immigration rate in the next four years is at least 50% larger than the maximum immigration rate in the previous four years; this year is defined as the year at which the positive labor demand shock took place. Second, the average population in the 4 years preceding the year of the shock is at least 10,000 workers. Third, indexing the year at which the shock took place as t, the total number of immigrants to the shocked region over the period $t-4$ to t must be at least 5,000 workers. The first criterium uses a discontinuity in immigration rates to identify regions experiencing a positive labor demand shock. The second and third criterium restrict the shocks of interest to those taking place in microregions that are sufficiently large and that typically drive a sufficiently large number of workers into the shocked region.

Once we have identified a microregion that experienced a positive labor demand shock according to the criteria outlined in the previous paragraph, we focus on the workers that migrated to the shocked region during a period encompassing the two years prior and the four years subsequent to the shock, and relate the timing of their migration decision to different characteristics of both the migrants themselves and of the microregion from which they migrated.

Given that the last year for which we observe migration decisions in our data is 2014 and that we want to study the evolution of the migration flows to a region experiencing a positive labor demand shock in the four years subsequent to the shock,

Table 1.1: Description of Positive Local Labor Demand Shocks

Microregion	State	Year Shock	Nb Migrants	Pop. Before	Shock Description
Ipojuca	Pernambuco	2009	89,067	48,119	Refinery Construction
Porto Velho	Rondônia	2008	74,352	121,808	Dam Construction
Altamira	Pará	2010	45,509	18,164	Dam Construction
Medio Mearim	Maranhão	2010	17,516	17,863	University Opening
Quirinópolis	Goiás	2006	11,939	12,488	Sugar Cane Energy

we only search for regions that experienced positive labor demand shocks prior to 2010. This procedure to identify such regions identified a total of five microregions. As summarized in Table B.4.1, these are diverse in their location, the year in which they took place, the total number of workers they draw into the shocked regions and the underlying cause of the shock.

For each of the labor demand shocks listed in Table B.4.1, we identify the workers that migrated to the shocked region in the two years prior or in the four years subsequent to the labor demand shock, and estimate the following regression model on this subset of workers:

$$d_{it+s} = \sum_{k=1}^{K} \sum_{s=-2}^{4} \gamma_s^k X_{it+s}^k + u_{it+s} \tag{1.1}$$

where t indicates the year in which the labor demand shock in the region of interest took place; i indexes a worker that migrated to the region of interest between $t-2$ and $t+4$; d_{it+s} is a dummy variable taking value 1 if i moved to the shocked region at year $t+s$; k indexes the covariates of interest, which we denote as $\{X_{it+s}^k; k = 1, \ldots, K\}$; u_{it+s} captures unobserved determinants of the migration dummy d_{it+s}; and, $\{\gamma_{it+s}^k; k = 1, \ldots, K\}$ is the parameter vector of interest. Denoting as $l(i, t+s-1)$ the microregion of residence of worker i at period $t+s-1$, and l' as the microregion experiencing the positive labor demand shock at period t, we include the following covariates X_{it+l}^k in our regression specification: a set of seven year dummies, each of

them taking value one for one of the years between $t-2$ and $t+4$; the log distance between $l(i,t+s-1)$ and l'; the per capita number of broadband internet connections, the population density, and the average annual wage in $l(i,t+s-1)$ at year $t+s-1$; the average migration share from $l(i,t+s-1)$ to l' between the years $t-5$ and $t-3$; the gender of worker i as well as her age and gender at period $t+s-1$; and, finally, a dummy variable taking value one if worker i was employed in the construction sector at period $t+s-1$. The reason for differentiating the construction workers in our regression specification is that, as illustrated in Table B.4.1, several of the labor demand shocks that we identify are linked to the construction of a large public infrastructure project in the microregion of interest l' and, thus, it would not be surprising to observe that construction workers moved relatively early to l'.

We think of the year dummies, the migration share from $l(i,t+s-1)$ to l' at $t+s-1$, and the dummy variable taking value one if worker i was employed in the construction sector at period $t+s-1$ as control variables that capture factors that both may explain why workers react to the labor demand shock of interest with different lags or delays and may be completely unrelated to heterogeneity in the information that workers may have about the labor demand shock taking place in region l' at period t. Conversely, we interpret all remaining covariates included in the regression specification described in equation (1.1) as capturing possible sources of heterogeneity in the information that worker i has about the labor demand shock of interest.

Figure 1.2 illustrates the estimates of $\{\gamma^k_{it+s}; k=1,\ldots,K\}$ that we obtain when focusing on the delay with which migrants reacted to the labor demand shock that took place in Altamira in 2010. Figures A.2.1 to A.2.4 in the Appendix contain analogous figures for all remaining cases listed in Table B.4.1. Figure 1.3 illustrates the estimates that arise from pooling the data across the five shock listed in Table B.4.1. All figures have the same structure. Each of the ten panels included in each figure

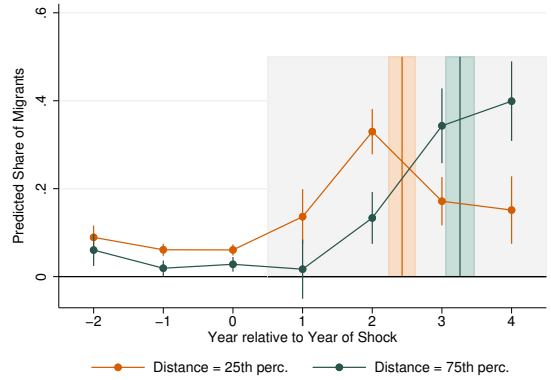
illustrates the importance of a different covariate X^k_{it+s}. Specifically, they illustrate, for each year between $t-2$ and $t+4$, the predicted migration probability for different values of the corresponding covariate X^k_{it+s}. For continuous covariates, we compare the predicted migration probability for X^k_{it+s} set to its 25% or its 75% percentile, with all other covariates set to their mean values. For categorical variables, the panel compares the predicted migration probabilities in each category. The whiskers attached to each dot represent the 95% confidence interval for each predicted migration probability. The dark thin vertical lines indicate the estimated expected number of years of delay implied by the expected probabilities, and the light-colored vertical bands illustrate the corresponding 95% confidence intervals. The standard errors of the estimates of the implied expected probabilities and the expected number of years of delay are computed using the Delta method and standard errors for the estimates of the parameter $\{\gamma^k_{it+s}; k = 1, \ldots, K\}$ clustered by year and microregion of origin of the migrant (i.e. by the microregion in which the migrant was located at period $t+s-1$).

The results in Figures 1.2 and 1.3 illustrate similar migration patterns. Everything else equal, workers living in geographically close microregions or in microregions with higher broadband internet penetration tend to react faster to positive labor demand shocks happening in microregions other than their location of residence. These relationships between both distance to the potential destination region and the degree of internet penetration feature also in the cases considered in Figures A.2.1 to A.2.4. Figure 1.2 also indicates that workers living in microregions with a higher average income, with larger migration past migration flows, and workers who identify as white, tend to react faster; however, these patterns is not as strong in the remaining labor demand shocks listed in Table B.4.1, and do not appear significant in Figure 1.3. Finally, it is the case for all shocks considered in our analysis that neither the education

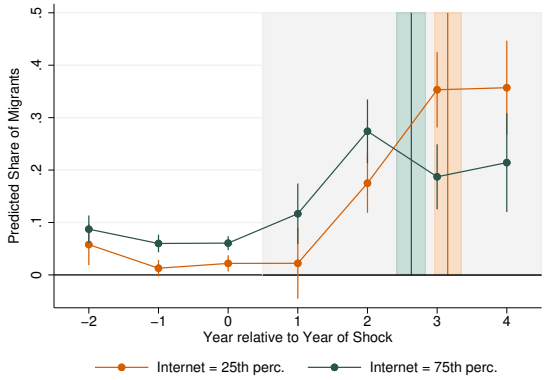
or gender of workers seem to affect the delay with which they react to labor demand shocks.

We interpret the estimates illustrated in Figures 1.2 and 1.3 and in Figures A.2.1 to A.2.4 as suggestive of the hypotheses that, everything else equal: (a) workers tend to have better information about labor demand shocks taking place in markets that are geographically close to their location of residence; and (b) workers located in areas with higher internet penetration have better information about every labor demand shocks, no matter where this one took place. In the remaining of this paper, we use an alternative identification strategy to test these hypotheses. Furthermore, we explore not only whether workers have heterogeneous information sets, but also try to document which variables are potentially useful to predict the wage that workers of a group would obtain upon migrating to a region are actually known by workers of these group.

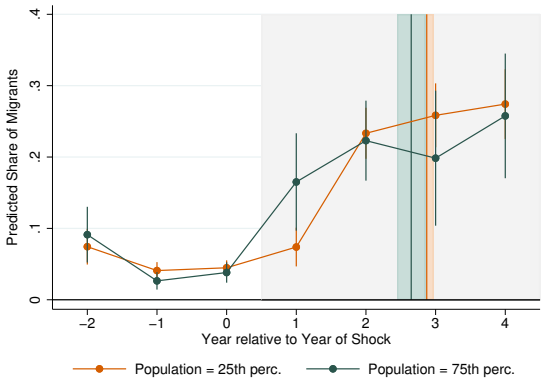
Figure 1.2: Heterogeneity in Migration Responses: Labor Market Shock in Altamira



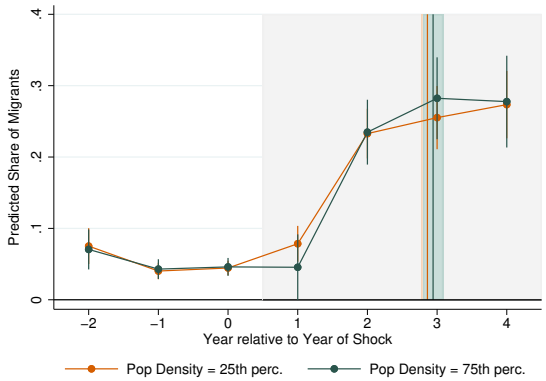
(a) Distance



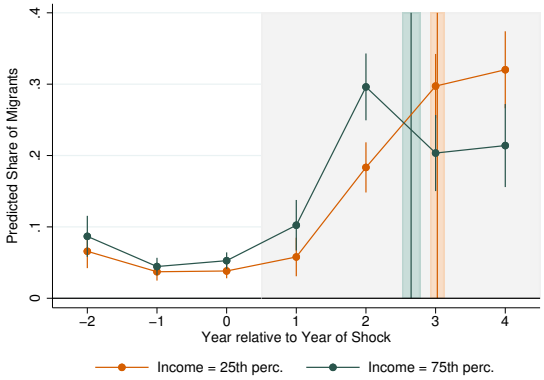
(b) Internet



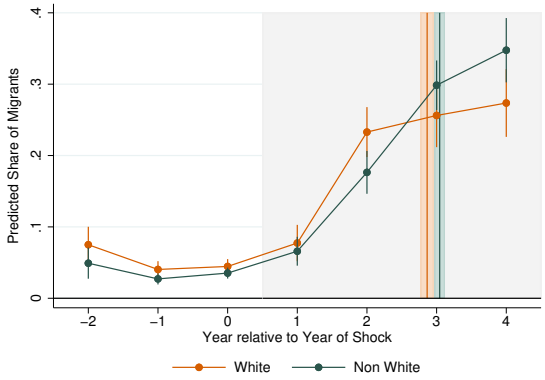
(c) Population



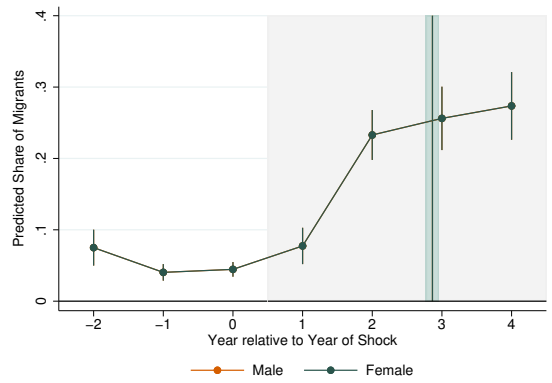
(d) Population Density



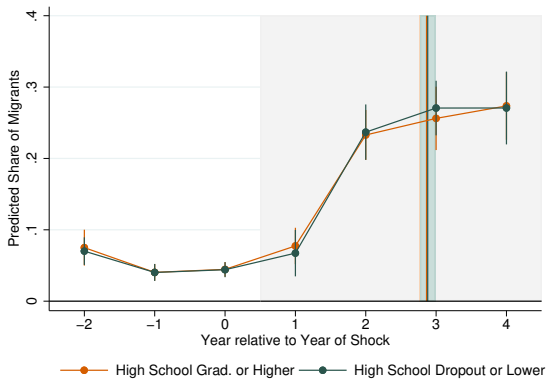
(e) Average Income



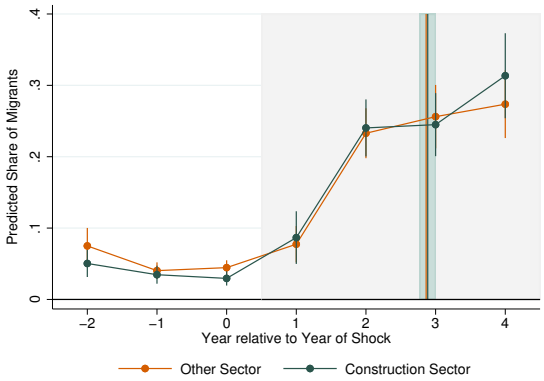
(f) Race



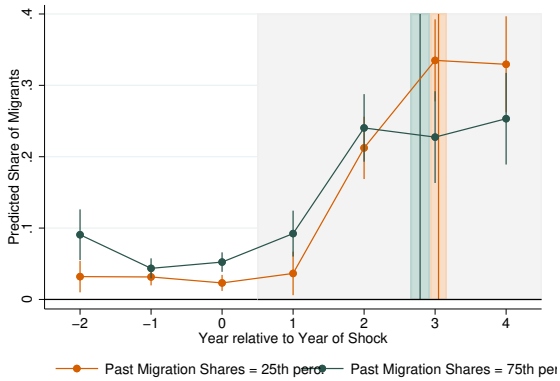
(g) Gender



(h) Education



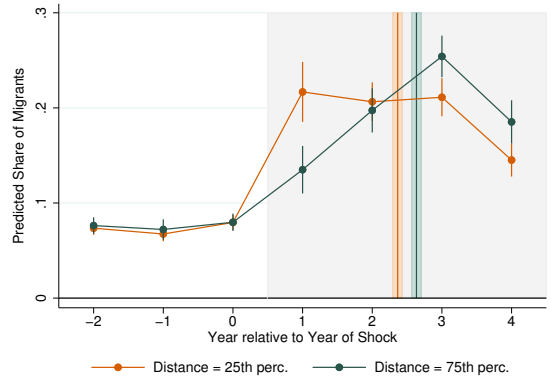
(i) Construction Sector



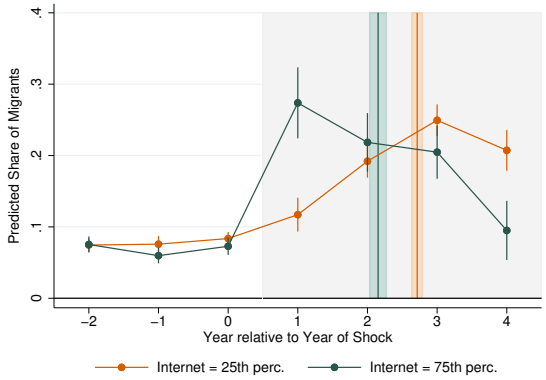
(j) Past Migration Flows

Note: Predicted migration probability to the shocked location, by year, for different values of the covariates. The whiskers attached to each dot represent the 95% confidence interval for each predicted migration probability. The dark thin vertical lines indicate the estimated expected number of years of delay implied by the expected probabilities, and the light-colored vertical bands illustrate the corresponding 95% confidence intervals. The standard errors of the estimates of the implied expected probabilities and the expected number of years of delay are computed using the Delta method and standard errors for the estimates of the parameter $\{\gamma^k_{it+s}; k = 1, \ldots, K\}$ clustered by year and microregion of origin of the migrant (i.e. by the microregion in which the migrant was located at period $t + s - 1$). Source: RAIS.

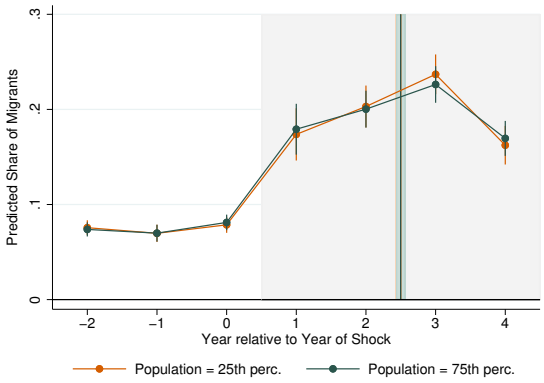
Figure 1.3: Heterogeneity in Migration Responses: Pooled Labor Market Shocks



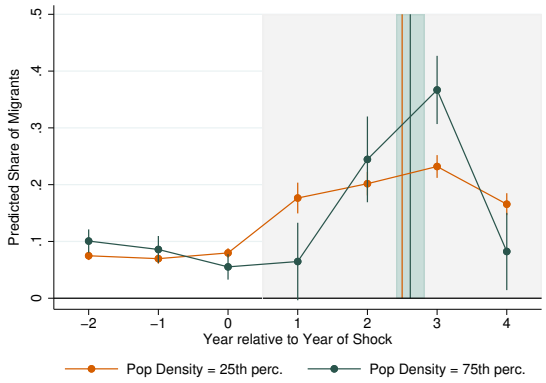
(a) Distance



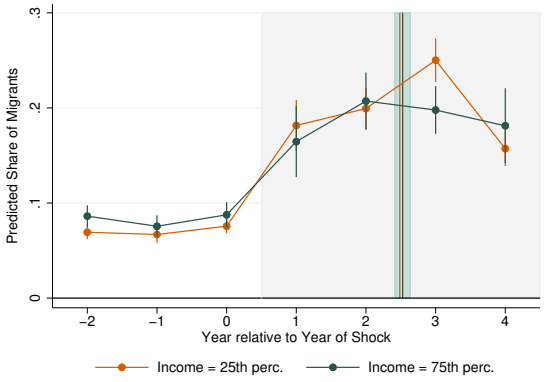
(b) Internet



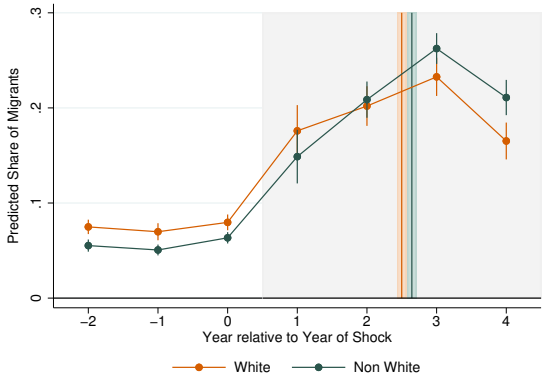
(c) Population



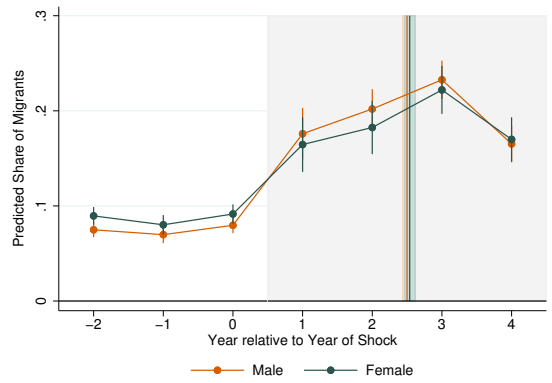
(d) Population Density



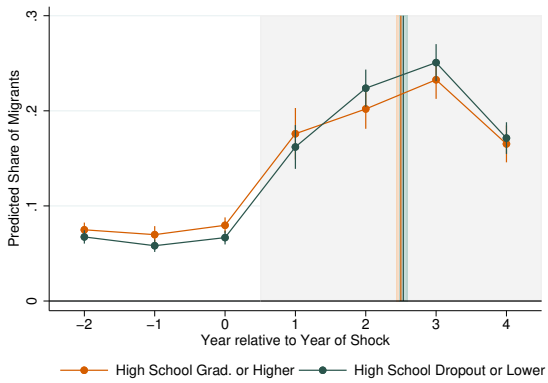
(e) Average Income



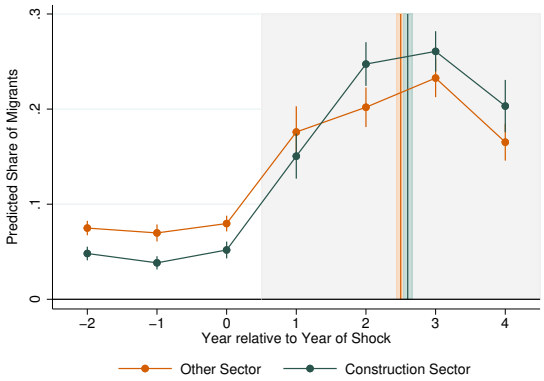
(f) Race



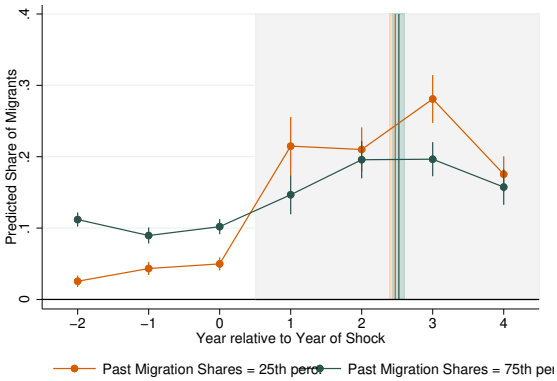
(g) Gender



(h) Education



(i) Construction Sector



(j) Past Migration Flows

Note: Predicted migration probability to the shocked location, by year, for different values of the covariates. The whiskers attached to each dot represent the 95% confidence interval for each predicted migration probability. The dark thin vertical lines indicate the estimated expected number of years of delay implied by the expected probabilities, and the light-colored vertical bands illustrate the corresponding 95% confidence intervals. The standard errors of the estimates of the implied expected probabilities and the expected number of years of delay are computed using the Delta method and standard errors for the estimates of the parameter $\{\gamma^k_{it+s}; k = 1, \ldots, K\}$ clustered by year and microregion of origin of the migrant (i.e. by the microregion in which the migrant was located at period $t + s - 1$). Source: RAIS.

1.3 Empirical Framework

Our approach to the identification of the content of migrants' information sets requires modeling how their migration decisions depend on their expectations of the current and future real wages and amenities prevalent in any location they may con-

sider migrating to. Our identification approach allows for the model-predicted mapping between workers' expectations and migration decisions to depend on unknown parameters; however, given available data (which does not include the workers' expectations), these parameters must be (point or partially) identified under only minimal content requirements on workers' information sets. Misspecifications of the mapping between workers' expectations and migration decisions may lead us to wrong conclusions on the content of workers' information sets and, as a consequence, to wrong conclusions on the quantitative importance of the information frictions that workers face. Consequently, our aim in building our empirical model is to provide a general framework linking workers' expectations to migration decisions subject to the constraint that it must allow us to use the data described in Section 1.2.1 to infer the content of workers' information sets.

We provide two frameworks under which we perform our analysis of the content of workers' information sets. First, we provide a framework under which workers' preference parameters and migration costs are point-identified. Second, we provide a framework under which workers' preference parameters and migration costs are only partially identified. The advantage of the point-identified model is that it allows for a computationally simpler approach to the characterization of workers' information sets. The advantage of the partially-identified model is that it allows to relax several of the model assumptions imposed in the point-identified model.

In Section 1.3.1, we describe model characteristics that will be present in all models discussed in this paper. In Section 1.3.2, we describe the additional assumptions that we impose in two models that, under these extra assumptions, are point-identified.[13]

[13]In Appendix A.1.3, we describe a weaker set of assumptions than those imposed in Section 1.3.2 that however guarantee that the model is partially identified.

1.3.1 General Model Features

Although our interest is in understanding how informational frictions affect workers' decisions over where to live and work, such migration decisions are usually determined jointly with workers' decisions over the sector of employment. We thus present a model in which workers' make decisions over both employment margins.

In any year $t = 1, 2, \ldots$, a finite set of workers $i = 1, \ldots, I(t)$ decide the market in which to supply labor. Each market is defined as the combination of a sector and a region. We index markets by $m \in M$, regions by $l \in L$ and sectors by $n \in N$. The set N includes an outside option $n = 0$ that accounts for informal employment, home employment and unemployment. Thus, $m = (n, l)$ and M is the Cartesian product of the sets L and N; i.e. $M = L \times N$. We denote the subset of markets in M that are located in a particular region l as M_l; i.e. $M_l \equiv \{(n', l'); (n', l') \in M, l' = l\}$. Analogously, we denote the subset of the markets of M that correspond to a particular sector n as M_n; i.e. $M_n \equiv \{(n', l'); (n', l') \in M, n' = n\}$. We also define functions $\check{l}(\cdot) : M \to L$ and $\check{n}(\cdot) : M \to N$ that respectively indicate the region and sector corresponding to market m; i.e. if $m = (n, l)$, then $\check{n}(m) = n$ and $\check{l}(m) = l$.

We model each worker's location as the outcome of a choice made by the corresponding worker. Every worker i decides in any year t her sector and region of residence as the outcome of a utility maximization problem defined over the set of markets M.[14] We use m_{it} to denote the optimal choice of market m for a worker i in period t. Analogously, we use n_{it} and l_{it} to denote the optimal choice of sector and region of worker i in period t; thus, $m_{it} = (n_{it}, l_{it})$ and it thus follows that $n_{it} = \check{n}(m_{it})$ and $l_{it} = \check{l}(m_{it})$.

[14] The model described in Appendix A.1.3 could be further relaxed to allow for consideration sets $M(i, t) \subseteq M$ that are heterogeneous across workers and time periods. We assume however for simplicity in the exposition that every workers' consideration set equals the complete set of possible markets.

The static utility of worker i choosing to work in market m at period t is a linearly additive function of several terms. First, for each market $m = (n, l)$, we consider region- and sector-specific amenity indices, which we denote as a_{lit} and b_{nit}, respectively. Second, we include a consumption index, which equals the ratio between the total income of the worker upon choosing market m, w_{mit}, and the price index in market m's location, p_{lit}.[15] Third, we allow for individual- and year-specific sectoral reallocation costs, $\gamma_{\bar{n}nit}$, which worker i faces if and only if the sector corresponding to market m differs from her previous sector of employment, which we denote as $\bar{n} = n_{it-1}$. Fourth, we let workers face individual- and year-specific migration costs, $\kappa_{\bar{l}lit}$, for worker i if the location corresponding to market m differs from her previous location, which we denote as $\bar{l} = l_{it-1}$. Finally, workers draw an individual-, market-, and year-specific preference shock ν_{mit}.

General Assumption 1 (Static Utility)**.** *We assume that the static utility of choosing market $m = (n, l)$ for a worker i in period t, who was previously in market $m_{it-1} = \bar{m} = (\bar{n}, \bar{l})$, is*

$$U_{\bar{m}mit} \equiv u_{\bar{m}mit} + \nu_{mit}, \tag{1.2a}$$

$$u_{\bar{m}mit} \equiv \alpha w_{mit} - \alpha p_{lit} - \kappa_{\bar{l}lit} - \gamma_{\bar{n}nit} + a_{lit} + b_{nit}, \tag{1.2b}$$

where

$$\gamma_{\bar{n}nit} = 0 \qquad \text{if } n = \bar{n}, \tag{1.3a}$$

$$\kappa_{\bar{l}lit} = 0 \qquad \text{if } l = \bar{l}, \tag{1.3b}$$

[15] Following Caliendo et al. (2019), we assume that workers have a logarithmic utility over their consumption index. Without loss of generality, we think of this consumption index as an aggregator over a set of good-specific consumption levels. See Matsuyama and Ushchev (2017) for a description of the conditions that this aggregator must satisfy so that there exists a common price aggregator that allows to represent a consumption index as the ratio of the worker's total income to this price aggregator.

and α is an unknown preference parameter.

According to equation (1.2a), the utility to worker i in period t of working in market m is a linearly additive function of the preference shock ν_{mit} and the remaining terms, as captured in the term $u_{\bar{m}mit}$. As indicated in equation (1.2b), this term $u_{\bar{m}mit}$ is itself the sum of the contribution of the nominal wage w_{mit} minus the location-specific price index and sector- and location-specific moving costs, to which are added sector- and location-specific amenities. Equation (1.2b) also indicates that, other than the preference shock ν_{mit}, the only component of utility that varies across both occupations and locations is the total per-period income w_{mit}. The parameter α determines the impact of log real wages (which equals log real consumption in our framework) on workers' utility.

The observed labor market of worker i at period t, m_{it}, is selected optimally by worker i at t. Specifically, worker i chooses at t the labor market m that maximizes her expected discounted inter-temporal utility among all the labor markets included in the set M. Workers' expectations are rational and, thus, equal the expectation with respect to the data generating process of the determinants of workers current and future utility levels conditional on their respective information sets. We allow the information set of worker i at period t, $\mathcal{V}_{it}$, to vary flexibly across workers and over time, which implies that workers' expectations may also differ flexibly across workers and time periods.

The discounted inter-temporal utility of worker i at period t upon choosing market m, initially in market $m_{it-1} = \bar{m}$, which we denote as $V_{\bar{m}mit}$, is the sum of the static utility associated to market m at period t, U_{mit}, and the discounted static utilities that worker i would earn over all future periods given the choice of market m at

period t.[16] We assume that the discount factor is the same for every firm and time period and denote it as δ.

We use $m_{it'|mt}$ to denote the choice that worker i would make at a period $t' > t$ given the choice of market m at period t. Analogously, we use $U_{\bar{m}'m'it'|mt}$ to denote the static utility that worker i would obtain at period $t' > t$ if she were to chose market m' in this time period, market $\bar{m}'$ in the period $t' - 1$, and she had chosen market m at period t. Formally, we define $V_{\bar{m}mit}$ as

$$V_{\bar{m}mit} \equiv u_{\bar{m}mit} + \delta \mathsf{V}_{mit+1} + \nu_{mit}. \tag{1.4}$$

The function V_{mit+1} denotes the value function of worker i at period $t+1$ conditional on having chosen option m at period t. It may be written as a function of future choices and the static utilities associated to these choices as

$$\mathsf{V}_{mit+1} \equiv \sum_{t'=t+1}^{\infty} \sum_{\bar{m}', m' \in M} \delta^{(t'-(t+1))} \mathbb{1}\{m' = m_{it'|mt}, \bar{m}' = m_{it'-1|mt}\} U_{\bar{m}'m'it'|mt} \tag{1.5}$$

where $\mathbb{1}\{A\}$ is an indicator function taking value 1 if A is true.

Given the definition of the worker's payoff function $V_{\bar{m}mit}$ in equations (1.4) and (1.5), the following assumption states precisely the restrictions we impose on workers' optimal choices.

General Assumption 2 (Optimal Choice of Labor Market)**.** *For every worker i and period t, who chose market $\bar{m}$ in period $t-1$,*

$$m_{it} = \underset{m \in M}{\operatorname{argmax}} \, \mathbb{E}\left[V_{\bar{m}mit} | \mathcal{V}_{it}\right], \tag{1.6}$$

[16] The model described in Appendix A.1.3 could be further relaxed to allow for planning horizons $T(i,t) < \infty$ that are heterogeneous across workers and time periods. We assume however for simplicity in the exposition that every workers' planning horizon includes every future year.

where $\mathbb{E}[\cdot]$ *denotes the expectation with respect to the data generating process and* $V_{\bar{m}mit}$ *is defined as in equations* (1.4) *and* (1.5).

As equations (1.4) and (1.5) illustrate, worker i's payoff function of choosing market m in year t, starting from market $\bar{m}$, $V_{\bar{m}mit}$, depends on: (a) the static utility at t, $U_{\bar{m}mit}$; and, (b) for every $t' > t$, the sequence of static utilities associated to the choices at later periods t' that would be optimal if worker i were to choose market m at period t. Thus, $\mathcal{V}_{it}$ incorporates any variable known by individual i at period t that helps her predict any of these determinants of $V_{\bar{m}mit}$.

We assume that, for every worker i and period t, the information set $\mathcal{V}_{it}$ includes the vector of idiosyncratic preferences $\nu_{it} \equiv (\nu_{mit}; m \in M)$.[17] We denote the remaining elements of the vector $\mathcal{V}_{it}$ as $\mathcal{W}_{it}$. Formally, we impose the following assumption.

General Assumption 3 (Information Set)**.** *For every worker* i *and period* t

$$\mathcal{V}_{it} = (\mathcal{W}_{it}, \nu_{it}). \tag{1.7}$$

This assumption imposes no restriction on the content of $\mathcal{W}_{it}$. In Section 1.3.2, we specify which restrictions must be imposed on the content of $\mathcal{W}_{it}$ in order to point-identify workers' preference parameters.

Generally, we can think of $\mathcal{W}_{it}$ as including, for every market $m \in M$, an m-specific vector $\mathcal{W}_{mit}$ that contains variables that help worker i at period t predict the value of $u_{\bar{m}mit'}$ for every $t' > t$; i.e. $\mathcal{W}_{it} \equiv (\mathcal{W}_{mit}; m \in M)$. In turn, we can think of each $\mathcal{W}_{mit}$ as incorporating a countable (possibly infinite) set $\mathcal{S}_{mit}$ of random variables W^s_{mit}; i.e. $\mathcal{W}_{mit} \equiv (W^s_{mit}; s \in \mathcal{S}_{mit})$. Under this description of the content of the information set $\mathcal{W}_{it}$, Assumption GA3 allows the sets $\mathcal{S}_{mit}$ to vary flexibly across

[17] For every index a, row vector A_a, and set $\mathcal{A}$, we use the notation $(A_a; a \in \mathcal{A})$ to denote a row vector that incorporates (in increasing order according of the index a) all row vectors A_a such that a belongs to $\mathcal{A}$. If A_a is a scalar (as ν_{mit} indeed is), then $(A_a; a \in \mathcal{A})$ is a row vector whose length equals the cardinality of the set $\mathcal{A}$.

markets, workers and time periods, implying that different workers at the same point in time as well as the same worker at different points in time may have very different information about the different labor markets $m \in M$. For example, if a worker i knows at period t the lagged price index of a region $l = \check{l}(m)$, then $p_{lit-1} \subseteq \mathcal{W}_{mit}$ or, equivalently, there exists an index s such that $s \in \mathcal{S}_{mit}$ and $W^s_{mit} = p_{lit-1}$. If such information is not available to worker i at period t for some location $l' = \check{l}(m')$, then $p_{l'it-1} \nsubseteq \mathcal{W}_{m'it}$ or, equivalently, there does not exist any index s included in $\mathcal{S}_{m'it}$ such that $W^s_{m'it} = p_{l'it-1}$.

As discussed in Section 1.2.1, although we observe at every period t the annual labor income of workers in the market in which they are effective employed, we naturally have no direct measure of the annual labor income that workers would have obtained at t if they had been employed in either a different sector or a different location than that in which they were effectively employed. However, in order to either point or partially identify workers' preference parameters, we need to construct a measure of w_{mit} for every market m, worker i and period t; we denote this measure as $\widehat{w}_{mit}$. The following assumption states the restriction that we impose on any possible difference between the actual wage that a worker i would have obtained in a market m if employed at period t and our measure of this wage.

General Assumption 4 (Measurement Error in Wages)**.** *For every market $m \in M$, worker i, and period t, we observe a wage measure $\widehat{w}_{mit}$ such that the measurement error $\varepsilon_{mit} \equiv \widehat{w}_{mit} - w_{mit}$ satisfies that $\mathbb{E}[\varepsilon_{mit}|\mathcal{W}_{it}] = 0$.*

We can thus rewrite $u_{\bar{m}mit}$ in equation (1.2b) as a function of our observed wage measure as

$$u_{\bar{m}mit} \equiv \alpha\widehat{w}_{mit} - \alpha p_{lit} - \kappa_{\bar{l}lit} - \gamma_{\bar{n}nit} + a_{lit} + b_{nit} - \alpha\varepsilon_{mit}. \tag{1.8}$$

Finally, we impose several restrictions on the distribution of the idiosyncratic preferences ν_{it}. To express these assumptions, it is convenient to define the following vector of payoff relevant variables

$$X_{it} \equiv (w_{mit}, p_{lit}, \kappa_{\bar{l}lit}, \gamma_{\bar{n}nit}, a_{lit}, b_{nit}), \tag{1.9}$$

where $m = m_{it}$, $n = n_{it}$, $l = l_{it}$, $\bar{n} = n_{it-1}$ and $\bar{l} = l_{it-1}$.

General Assumption 5 (Distribution of Idiosyncratic Preferences)**.** *The vector ν_{it} satisfies the following restrictions:*

1. $\nu_{it} \perp\!\!\!\perp \nu_{i't'}$ *if* $i \neq i'$ *or* $t \neq t'$*;*

2. $\nu_{it} \perp\!\!\!\perp (\mathcal{W}_{i't'}, X_{i't'}, \varepsilon_{i't'})$ *for any* i*,* i'*,* t*, and* t'*.*

3. *The cumulative distribution function of the vector ν_{it} is:*

$$F(\nu_{it}) = \exp\left(-\sum_{n \in N} \Big(\sum_{m \in M_n} \exp(-\nu_{mit}/\lambda) \Big)^{\lambda} \right), \tag{1.10}$$

for $\lambda \in (0, 1]$.

The first restriction in Assumption GA5 imposes that that the vector of idiosyncratic utility shocks ν_{it} is independent across individuals i and time periods t. The second restriction imposes that the vector ν_{it} is independent of the remaining elements of the information set $\mathcal{V}_{i't'}$—see equation (1.7)—, of the vector of payoff relevant variables $X_{i't'}$, and of the vector of measurement errors in our wage measures, $\varepsilon_{i't'}$, for any pair of workers i and i' and any pair of periods t and t'. The third restriction in Assumption GA5 imposes that the distribution of the vector ν_{it} is a type of Generalized Extreme Value distribution that satisfies the following restrictions: (a) the

marginal distribution of each ν_{mit} is univariate extreme value; (b) any two idiosyncratic preference shocks ν_{mit} and $\nu_{m'it}$ such that $\check{n}(m) \neq \check{n}(m')$ are independent; and (c) any two idiosyncratic preference shocks ν_{mit} and $\nu_{m'it}$ such that $\check{n}(m) = \check{n}(m')$ may be independent (if $\lambda = 1$) or dependent (if $\lambda \in (0, 1)$).

1.3.2 Limited Heterogeneity Model: Group-Specific Information Sets

The restrictions imposed in Assumptions GA1 to GA5 are not sufficient to identify the information that different workers have about the payoff of choosing different labor markets. We thus impose here several additional restrictions to those imposed in Assumptions GA1 to GA5.

A first set of additional restrictions serve the purpose of limiting the heterogeneity in the vector of payoff relevant variables defined in equation (1.9). The framework described in Section 1.3.1 allows sector-specific and location-specific amenities, location-specific price indices, and market-specific wages, to vary flexibly across workers i and time periods t. While we allow these payoff relevant variables to vary flexibly across time periods, we impose here the restriction that, in any given period, these payoff relevant variables are common across workers of the same demographic type. We index demographic types by $d = 1, \ldots, D$, use d_i to denote the demographic type of worker i, and assume that any two workers are of the same demographic type if they have the same gender, race, and education level. Formally, we impose:

Limited Heterogeneity Assumption 1 (Amenities, Price Indices, and Wages)**.** *For any period t, any market $m = (n, l)$, and any two workers i and i' such that $d_i = d_{i'}$, it holds that:*

$$b_{nit} = b_{ni't}$$

$$a_{lit} = a_{li't}$$

$$p_{lit} = p_{li't}$$

$$w_{mit} = w_{mi't}.$$

Thus, for any worker i in demographic type $d = d_i$, period t and market m, we can write b_{nit}, a_{lit}, p_{lit}, and w_{mit} as b_{ndt}, a_{ldt}, p_{ldt}, and w_{mdt}, respectively. We impose no additional restriction on how amenities and price indices vary across markets, demographic types and periods.

We also impose similar assumptions limiting the heterogeneity in sectoral switching costs and migration costs. Specifically, we assume that, in any given period, sectoral switching costs are common to workers who are of the same demographic type and were employed in the same sector in the previous period. Similarly, we assume that, in any given period, migration costs are common to workers who are of the same demographic type and were employed in the same location in the previous period. As a reminder, we denote the prior occupation of a worker i at period t as $\bar{n} = n_{it-1}$ and its previous location as $\bar{l} = l_{it-1}$. Using this notation, we state the restrictions on occupation switching costs and migration costs formally in the next two assumptions.

Limited Heterogeneity Assumption 2 (Occupation Switching Costs)**.** *For any period t, any market $m = (n, l)$, and any two workers i and i' such that $d_i = d_{i'}$ and $\bar{n} = n_{it-1} = n_{i't-1}$, it holds that:*

$$\gamma_{\bar{n}nit} = \gamma_{\bar{n}ni't}.$$

Limited Heterogeneity Assumption 3 (Migration Costs)**.** *For any period t, any market $m = (n, l)$, and any two workers i and i' such that $d_i = d_{i'}$ and $\bar{l} = l_{it-1} =$*

$l_{i't-1}$, *it holds that:*

$$\kappa_{\bar{l}\bar{l}it} = \kappa_{\bar{l}\bar{l}i't}.$$

Thus, for any worker i in demographic type $d = d_i$, period t and market $m = (n, l)$, we can write $\gamma_{\bar{n}nit}$ and $\kappa_{\bar{l}\bar{l}it}$ as $\gamma_{\bar{n}ndt}$ and $\kappa_{\bar{l}\bar{l}dt}$. We impose no additional restriction on how occupation-specific switching costs and migration costs vary across markets, demographic types, workers' prior sectors and locations, and periods.

We similarly limit the heterogeneity of workers' information sets. Specifically, we assume that, in any given period, information sets are common to workers that are of the same demographic type and worked in the same market (sector and location) in the previous period. We denote the prior market of employment of a worker i at period t (i.e. its market at period $t-1$) as $\bar{m} = m_{it-1}$. We denote workers that are of the same demographic type and worked in the same market in the previous period as belonging to the same group, index worker groups by $g = 1, \ldots, G$, and denote the group to which a worker i belongs at period t as g_{it}; i.e. $g_{it} = g_{i't'}$ if $d_i = d_{i'}$ and $m_{it-1} = m_{i't'-1}$. We formally state the restrictions on the information sets in the next assumption.

Limited Heterogeneity Assumption 4 (Information Set)**.** *For any period t and any two workers i and i' such that $g_{it} = g_{i't}$, it holds that*

$$\mathcal{W}_{it} = \mathcal{W}_{i't}.$$

Thus, for any worker i and period t, such that worker i is in group $g = (\bar{m}, d)$, i.e. $g_{it} = (m_{it-1}, d_i)$, we can write $\mathcal{W}_{it}$ as $\mathcal{W}_{gt}$.[18] We impose no restriction on how information sets vary across worker groups or time periods.

Importantly, as implied by Assumptions LH1 to LH4, in any given period t, any two workers that belong to the same group will share not only the same information set but they will also share: the same amenities in every sector and location, the same price index in every location, and the same wages in every market (as implied by Assumption LH1); the same switching costs to any sector (as implied by Assumption LH2); and, the same migration costs to any destination (as implied by Assumption LH3). Therefore, we can refer as $d(g)$ the demographic type of those workers included in group g, and as $\bar{n}(g)$ and $\bar{l}(g)$ the previous sector and work locations of those workers included in group g. We denote as $m_t(g)$, $n_t(g)$ and $l_t(g)$ the optimal market, sector and location of employment chosen by a worker of type g at period t.

Finally, to complete the model in such a way that one may point-identify the workers' preference parameters, one needs to specify for every worker group g and period t a vector Z^*_{gt} such that Z^*_{gt} belongs to the information set $\mathcal{W}_{gt}$. Formally, we refer to a vector Z^*_{gt} as belonging to the information set $\mathcal{W}_{gt}$ if the distribution of Z^*_{gt} conditional on $\mathcal{W}_{gt}$ is degenerate, and use the notation $Z^*_{gt} \subseteq \mathcal{W}_{gt}$ to describe this relationship between the random vectors Z^*_{gt} and $\mathcal{W}_{gt}$.

Identifying a vector Z^*_{gt} that is known by workers of type g in period t is not trivial, as it requires specifying a covariate that every worker in a particular group and year knows. We assume that, for every worker group g, any period t, and any microregion l, workers belonging to group g at period t know whether the average annual income in period $t-1$ and sector $\bar{n}(g)$ for workers of type $d(g)$ was larger or

[18] The lack of heterogeneity in information sets across workers that belong to the same group (i.e. across workers of the same demographic type, in the same year, and that were employed in the same market in the prior year) is the main restriction that the point-identified model described in this section imposes, and that the partially-identified model described in Appendix A.1.3 does not.

smaller in the mesoregion where workers of group g were located at $t-1$ than in the mesoregion to which microregion l belongs. Formally, we can define Z^*_{gt} as

$$Z^*_{gt} = (Z^*_{lgt}; l = 1, \ldots, L), \tag{1.11}$$

where Z^*_{lgt}, with $g = (\bar{m}, d)$ and $\bar{m} = (\bar{n}, \bar{l})$, is a dummy variable that takes value one if the average annual income in period $t-1$ and sector $\bar{n}$ for workers of type d was larger in mesoregion to which location l belongs, which we denote as s, than in mesoregion to which location $\bar{l}$ belongs, which we denote as $\bar{s}$; i.e.

$$Z^*_{lgt} \equiv \mathbb{1}\left\{\overline{w}_{sd\bar{n}t-1} > \overline{w}_{\bar{s}d\bar{n}t-1}\right\} \tag{1.12}$$

and, for every mesoregion $s = 1, \ldots, S$, group $g = (\bar{m}, d)$, with $\bar{m} = (\bar{n}, \bar{l})$,

$$\overline{w}_{sd\bar{n}t-1} \equiv \frac{\sum_{m=1}^{M} \mathbb{1}\{s(\check{l}(m)) = s, \check{n}(m) = \bar{n}\} w_{mdt-1}}{\sum_{m=1}^{M} \mathbb{1}\{s(\check{l}(m)) = s, \check{n}(m) = \bar{n}\}}. \tag{1.13}$$

Limited Heterogeneity Assumption 5 (Minimal Content of Information Set)**.** *We assume that, for any group g and year t, $Z^*_{gt} \subseteq \mathcal{W}_{gt}$, with Z^*_{gt} defined as in equations* (1.11)*,* (1.12) *and* (1.13)*.*

Given that, according to Assumption LH1, all workers of the same demographic type d employed in the same market m at a given period t earn the same annual wage, we also construct a wage proxy that, given a market and a period, does not vary across workers of the same demographic type. We denote this proxy as $\widehat{w}_{mdt}$, and the measurement error affecting this proxy as ε_{mdt}; i.e.

$$\varepsilon_{mdt} \equiv \widehat{w}_{mdt} - w_{mdt}. \tag{1.14}$$

Assumption GA4 thus imposes that

$$\mathbb{E}[\varepsilon_{md(g)t}|\mathcal{W}_{gt}] = 0, \tag{1.15}$$

and, given this mean independence condition and the Law of Iterated Expectations, Assumption LH5 implies that

$$\mathbb{E}[\varepsilon_{md(g)t}|Z^*_{gt}] = 0. \tag{1.16}$$

1.4 Estimation of Migration Elasticity

We focus for identification on the decision of where to work taken by those workers that decide to remain employed in the same sector in which they were employed in the prior year.

In any given period t, the probability that a worker in group $g = (\bar{m}, d)$, with $\bar{m} = (\bar{n}, \bar{l})$, decides to remain employed in sector $n = \bar{n}$ and to work in any given location l, so that she chooses $m_t(g) = m = (\bar{n}, l)$, is

$$P(m = m_t(g)|\mathcal{W}_{it}, n = \bar{n}) = \frac{\exp\left((1/\lambda)\mathbb{E}[v_{gmt}|\mathcal{W}_{gt}]\right)}{\sum_{m' \in M_{\bar{n}(g)}} \exp\left((1/\lambda)\mathbb{E}[v_{gm't}|\mathcal{W}_{gt}]\right)}, \tag{1.17}$$

with v_{mgt} defined as the value of choosing market m in period t for an individual initially in group g,

$$v_{gmt} = \alpha w_{mdt} - \alpha p_{ldt} + a_{ldt} + b_{ndt} - \kappa_{\bar{l}ldt} - \gamma_{\bar{n}ndt} + \delta \mathsf{V}_{mdt+1}, \tag{1.18}$$

where V_{mdt+1}, defined in (1.5), denotes the value function of a worker of demographic type d at period $t+1$ conditional on having chosen market m at period t. In the following, we denote by $\mathfrak{g} = (m, d)$ the group of a worker that belonged to group g

at period t and chose market m at this period; i.e. $d(g) = d(\mathfrak{g})$, $\bar{n}(\mathfrak{g}) = \check{n}(m)$, and $\bar{l}(\mathfrak{g}) = \check{l}(m)$.

In this section, we maintain the assumption that $\lambda = 1$ but allow $\delta > 0$. As we show in Appendix A.1.2, under the assumptions that $\lambda = 1$ and $\delta = 0$, the model described in Sections 1.3.1 and 1.3.2 imply the following moment condition

$$\begin{aligned}\mathbb{E}\left[\ln\left(\frac{P(m = m_t(g)|\mathcal{W}_{gt})}{P(\bar{m} = m_t(g)|\mathcal{W}_{gt})}\left(\frac{P(m = m_{t+1}(\mathfrak{g})|\mathcal{W}_{\mathfrak{g}t+1})}{P(m = m_{t+1}(g)|\mathcal{W}_{gt+1})}\right)^{\delta}\right) \times Z^*_{lgt}\right] \\ = \mathbb{E}\left[\left(\chi_{\bar{l}ldt} + \alpha\left(\widehat{w}_{mdt} - \widehat{w}_{\bar{m}dt}\right)\right) \times Z^*_{lgt}\right], \qquad (1.19)\end{aligned}$$

where $\mathfrak{g} = (m, d)$ denotes the group to which workers that belonged to group g at period t will belong at period $t+1$ if they chose market m at period t. Note that the group to which workers that belonged to group g at period t will belong at period $t+1$ if they chose market $\bar{m}$ at period t is $g = (\bar{m}, d)$. In Section 1.6, we use this moment condition and the definition of Z^*_{lgt} for any location $l = 1, \ldots, L$ in equations (1.12) and (1.13) to present estimates of the elasticity of migration with respect to wages, α.

1.5 Identification of Information Frictions

The moment condition in equation (1.19) point-identify the parameter capturing the elasticity of migration flows with respect to wages, α. However, one could construct analogous moment conditions substituting the variable Z^*_{lgt} in equation (1.19) by any other variable belonging to the information set of workers of group g at period t. We index these other potential variables by $k = 1, \ldots, K$, and denote them as $Z_{k,lgt}$.

If it is true that $Z_{k,lgt} \subseteq \mathcal{W}_{gt}$, then moment conditions analogous to those in equation (1.19), but that use $Z_{k,lgt}$ instead of Z^*_{lgt} should also hold at the true value of

the parameter vector α. On the contrary, if $Z_{k,lgt} \not\subseteq \mathcal{W}_{gt}$, then the moment conditions analogous to those in equation (1.19), but that use $Z_{k,lgt}$ instead of Z^*_{lgt} may not hold at the true value of the parameter vector. We combine the moment conditions in equation (1.19) with analogous moment conditions that use alternative variables $Z_{k,lgt}$ in the place of Z^*_{lgt}, and implement a J-test of overidentifying restrictions to determine whether all moment conditions are jointly valid. If we fail to reject the null that all moment conditions we test are jointly valid, then we conclude that we cannot reject that $Z_{k,lgt}$ belongs to the information set of workers belonging to group g at period t. If, on the contrary, we reject the null that all moment conditions are jointly valid, then we conclude that $Z_{k,lgt}$ does not belong to the information set of workers belonging to group g at period t.

This procedure allows us to test not only whether a particular variable $Z_{k,lgt}$ belongs to the information set $\mathcal{W}_{gt}$ for all worker groups g and all time periods t, but also whether $Z_{k,lgt}$ belongs to $\mathcal{W}_{gt}$ only for a subset of groups g, time periods t, and destination markets m. To do so, we build a moment condition that uses an instrument that equals $Z_{k,lgt}$ for that subset of groups, time periods and destination markets for which we posit that $Z_{k,lgt}$ belongs to the information set, and that uses an instrument that equals Z^*_{lgt} for all remaining groups, time periods and destination markets. As indicated above, once we combine such a moment condition with one that uses an instrument that equals Z^*_{lgt} for all groups, periods and markets (which, according to the model described in Sections 1.3.1 and 1.3.2, is valid), we can use a J-test of overidentifying restrictions to determine whether $Z_{k,lgt}$ belongs to $\mathcal{W}_{gt}$ for a subset of groups, time periods and destination markets.

In Section 1.6, we present several tests of overidentifying restrictions following the methodology described in this section.

1.6 Results

1.6.1 Estimates of Elasticity of Migration With Respect to Wages

In Table 1.2, we present estimates of the elasticity of migration flows with respect to wages, α. We base our estimates on moment conditions that exploit the moment function in equation (1.19), controlling for demographic type-origin-destination-year fixed effects.

The different columns of Table 1.2 combine the corresponding moment condition with different instruments in order to build the moment conditions used to estimate α. The first column uses the wage differential that enters the moment function,

$$\widehat{w}_{mdt} - \widehat{w}_{\bar{m}dt},$$

as the instrumental variable of the different moment conditions. This yields the Ordinary Least Squares (OLS) estimate of α. The elasticity of migration flows with respect to wages is 0.063.

The second column of Table 1.2 uses as instrument the one-period lag of the wage differential entering the moment function, with $g = (\bar{m}, d)$, $\bar{m} = (\bar{n}, \bar{l})$ and $m = (\bar{n}, l)$ for any $l = 1, \ldots, L$,

$$Z_{1,gt} = (Z_{1,lgt}; l = 1, \ldots, L), \tag{1.20a}$$

$$Z_{1,lgt} = \widehat{w}_{mdt-1} - \widehat{w}_{\bar{m}dt-1}, \tag{1.20b}$$

and the third column uses as instrument a dummy variable that takes value 1 if the one-period lag of the wage differential entering the moment function is positive,

$$Z_{2,gt} = (Z_{2,lgt}; l = 1, \ldots, L), \tag{1.21a}$$

$$Z_{2,lgt} = \mathbb{1}\{\widehat{w}_{mdt-1} > \widehat{w}_{\bar{m}dt-1}\}, \tag{1.21b}$$

Note that $Z_{1,gt}$ will be a valid instrument if workers in group $g = (\bar{m}, d)$, with $\bar{m} = (\bar{n}, \bar{l})$, know exactly the $t-1$ wage of workers of their same demographic type, d, in sector $\bar{n}$ in every Brazilian microregion. Similarly, $Z_{2,gt}$ will be a valid instrument if workers in group g at least know whether the $t-1$ wage of workers of demographic type d in sector $\bar{n}$ in every Brazilian microregion is larger or smaller than the corresponding wage in their microregion of residence at period $t-1$, $\bar{l}$. The informational requirements that the validity of the instrument $Z_{1,gt}$ imposes are not much weaker than those needed to assure the consistency of the OLS estimator reported in column 1 of Table 1.2. Yet, simply replacing the assumption that individuals know the current wage differences by the assumption that they know these differences at $t-1$ yields an estimates α of 0.181 in column 2 of Table 1.2 that 287% larger. In comparison, the informational requirement imposed by the validity of the instrument $Z_{2,gt}$ is significantly weaker since it only assumes that individuals are able to identify whether any $t-1$ wage is higher or lower than the local wage. Using $Z_{2,gt}$ as instrument, we obtain a higher migration elasticity of 0.252 in column 3, which is 4 times larger than the OLS estimates.

The fourth and fifth columns of Table 1.2 use as instruments variables that are analogous to those used as instruments in the second and third column, with the only exception that average wages are defined as the mesoregion to which each the corresponding microregion l belongs. Specifically, the instrumental variable used in

the fourth column is defined as

$$Z_{3,gt} = (Z_{3,lgt}; l = 1, \ldots, L), \tag{1.22a}$$

$$Z_{3,lgt} = \overline{w}_{sd\bar{n}t-1} - \overline{w}_{\bar{s}d\bar{n}t-1}, \tag{1.22b}$$

with $g = (\bar{m}, d)$ and $\bar{m} = (\bar{n}, \bar{l})$, where s and $\bar{s}$ denote the mesoregions in which the microregions l and $\bar{l}$ are located, respectively, and $\overline{w}_{sd\bar{n}t-1}$ is defined in (1.13).

The instrumental variable used in the fifth column is defined as in equations (1.11), (1.12) and (1.13). The validity of the instruments used in columns 4 and 5 requires workers to have information on wages at the mesoregion level (instead of having wage information at the microregion level). This naturally imposes weaker informational requirements on the part of the workers. As a consequence, the resulting estimates of α reported in columns 4 and 5 are larger than those reported in columns 1 to 3. Specifically, in comparison to the OLS estimate reported in column 1, the IV estimate reported in column 5 is approximately 813% larger.

The results in Table 1.2 illustrate that standard estimates of the elasticity of migration flows with respect to wage differentials are very sensitive to the assumptions that the researcher imposes on the content of the potential migrants' information sets. Specifically, our results suggest that estimates obtained following the standard procedure in the literature (which correspond to those reported in column 2) may suffer from a significant downward bias.

Table 1.2: Estimates of the Elasticity of Migration With Respect to Wages by OLS and with Different Instrumental Variables

	OLS	TSLS	TSLS	TSLS	TSLS
Instrumental Variable:		$Z_{1,gt}$	$Z_{2,gt}$	$Z_{3,gt}$	Z^*_{gt}
$\hat{\alpha}$	0.0630^a	0.181^a	0.252^a	0.303^c	0.512^a
	(0.0171)	(0.038)	(0.045)	(0.158)	(0.090)
Observations	191,911	191,911	191,911	191,911	191,911
First-stage IV Estimate		0.828^a	0.368^a	0.0184^a	0.0994^a
		(0.001)	(0.001)	(0.0004)	(0.0015)
First-stage Within R^2		0.68	0.27	0.0092	0.023

Note: a indicates significance at the 1% level, b at the 5% level and c at the 10% level. In parenthesis, we report two-way cluster robust standard errors that allow for correlation in regression residuals at the demographic type-origin-year level and at the demographic type-destination-year level.

1.6.2 Testing the Content of Information Sets

All Regions and Demographic Types

In Table 1.3, we follow the approach described in Section 1.5 to test whether certain variables belong to the information set of workers of all groups g in certain time periods t.

Adopting a 5% critical value, we observe that we can reject that workers of all groups in all time periods know any of the four possible variables we use as candidate instruments (see the definition of the variables X and $Z_{k,gt}$, $k = 1, 2, 3$, in Section 1.6.1). The estimates α in the last column are different from the ones reported in

Section 1.6.1 since they correspond to using the pair $(Z^*_{gt}, Z_{k,gt})$, $k = 1, 2, 3$, or (Z^*_{gt}, X) as instruments, instead of using $Z_{k,gt}$ or X as individual instruments in Table 1.2. The results in Table 1.3 may mask heterogeneity in individuals' information sets, either across groups (i.e. across demographic types or origin location) or across potential destinations.

Heterogeneity Across Destinations

We now explore whether the composition of individuals' information sets differs by group or destination. In Table 1.4, we look at whether individuals possess richer information sets about microregions that are located at a smaller geographical, racial or educational distance, that have a higher population density or that benefit from a higher rate of internet penetration.

We define the racial and educational distances between two locations in a given year as the absolute value of the difference between the two locations' share of white individuals and high school graduates or above, respectively. The median educational distance between microregions in the sample across years is 8.3 percentage points. Over the same period, the median racial distance is 8.4 percentage points and the median geographical distance is 147km. As defined in Section 2.3, the measure of internet access is the number of broadband connections per capita. The median number of connections per capita equals 10.8% over the period between 2007 and 2014, the years for which we have data on internet connections. The median population density at the microregion level is 59.7 inhabitants per square kilometer across all years.

In each of the panels reported in Table 1.4, the left column indicates the selecting criterion for the observation to be tested. For example, in the first panel, we define the instrument $\tilde{Z}_{k,gt}$, with $k = \{1, 2, 3\}$, which is equal to $Z_{k,gt}$ for observations corresponding to destinations located at least as far as 382 km from the origin (the

Table 1.3: Testing Content of Information Set - All Groups

Regions tested	Variable Tested	J-test statistic	J-test p-value	$\hat{\alpha}$
All	Z^*_{gt}	0	NaN	0.512
	$Z_{3,gt}$	2.98	0.0426	0.468
	$Z_{2,gt}$	7.07	0.0078	0.252
	$Z_{1,gt}$	10.91	0.0009	0.130
	X	8.57	0.0034	0.089

Note: In orange, the tests for which the p-value is below 0.05, indicating that we can reject the exogeneity of the tested variable with 95% confidence. The J-statistic is computed from the 2-step GMM estimation of α, using $(Z^*_{gt}, Z_{k,gt})$, $k = 1, 2, 3$, or (Z^*_{gt}, X) as instruments, and allowing for two-way clustering at the origin and destination microregions. The p-value is the probability that a random variable with a $\chi^2(1)$ distribution takes a value above J.

75th percentile of distance in the sample), and equals the reference instrument Z^*_{gt} for all other observations. Denoting by W the variable along which the selection criterion is defined, and remembering that observations vary at the group-year-destination level glt, where $g = (\bar{m}, d)$, $\bar{m} = (\bar{n}, \bar{l})$, we construct the following instruments

$$\tilde{Z}^+_{k,glt} \equiv \mathbb{1}\{W_{glt} > q^W_{75}\} Z_{k,glt} + \mathbb{1}\{W_{glt} \leq q^W_{75}\} Z^*_{glt},$$

$$\tilde{Z}^-_{k,glt} \equiv \mathbb{1}\{W_{glt} < q^W_{25}\} Z_{k,glt} + \mathbb{1}\{W_{glt} \geq q^W_{25}\} Z^*_{glt},$$

where $W_{glt} \in \{dist_{\bar{l}l}, internet_{lt}, popdensity_{lt}, educdist_{\bar{l}lt}, racedist_{\bar{l}lt}\}$ and q^W_{75}, q^W_{25}, denote the 75th and 25th percentile of the distribution of W in the sample, respectively. Defined this way, assuming the validity of $\tilde{Z}^+_{k,glt}$ and $\tilde{Z}^-_{k,glt}$ as instruments corresponds to imposing the assumption that the variable $\tilde{Z}_{k,gt}$ belongs to individuals' information sets only for the 25 percent of destinations selected by the criterion, while maintaining the initial assumption that the reference instrument is known about other destinations.

Table 1.4: Testing Content of Information Set - Heterogeneity Across Destinations

Regions Tested	Variable Tested	J-test Statistic	J-test p-value	$\hat{\alpha}$
Geographically far away Distance $>$ 382 km	$Z_{3,gt}$	5.24	0.0221	0.462
	$Z_{2,gt}$	11.25	0.0007	0.301
	$Z_{1,gt}$	12.06	0.0005	0.201
	X	7.05	0.0079	0.231
Geographically close Distance $<$ 68 km	$Z_{3,gt}$	0.16	0.6816	0.532
	$Z_{2,gt}$	1.92	0.1652	0.403
	$Z_{1,gt}$	7.90	0.0049	0.252
	X	11.05	0.0008	0.045
Low internet access at destination Sh. Internet $<$ 6.4 %	$Z_{3,gt}$	4.23	0.0397	0.421
	$Z_{2,gt}$	13.67	0.0002	0.143
	$Z_{1,gt}$	12.45	0.0004	0.111
	X	10.55	0.0011	0.121
High internet access at destination Sh. Internet $>$ 16 %	$Z_{3,gt}$	0.42	0.5122	0.504
	$Z_{2,gt}$	0.04	0.8285	0.514
	$Z_{1,gt}$	2.69	0.1006	0.494
	X	8.28	0.0040	0.160
Low population density at destination Pop. dens $<$ 11 inh/km^2	$Z_{3,gt}$	4.54	0.0331	0.452
	$Z_{2,gt}$	11.62	0.0006	0.179
	$Z_{1,gt}$	13.03	0.0003	0.100
	X	12.30	0.0004	0.054
High population density at destination Pop. dens $>$ 189 inh/km^2	$Z_{3,gt}$	1.15	0.2828	0.548
	$Z_{2,gt}$	5.69	0.0169	0.354

	$Z_{1,gt}$	8.81	0.0029	0.282
	X	7.48	0.0062	0.209
High Educational Distance $\|\Delta$ Sh. High. Sch.$\| > 14.4$ p.p.	$Z_{3,gt}$	4.69	0.0303	0.460
	$Z_{2,gt}$	11.52	0.0006	0.159
	$Z_{1,gt}$	12.51	0.0004	0.046
	X	9.91	0.0016	0.108
Low Educational Distance $\|\Delta$ Sh. High. Sch.$\| < 3.9$ p.p.	$Z_{3,gt}$	1.19	0.2752	0.570
	$Z_{2,gt}$	0.35	0.5487	0.472
	$Z_{1,gt}$	4.02	0.0448	0.361
	X	7.72	0.0054	0.142
High Racial Distance $\|\Delta$ Sh. White$\| > 16.3$ p.p.	$Z_{3,gt}$	5.30	0.0212	0.483
	$Z_{2,gt}$	11.40	0.0007	0.224
	$Z_{1,gt}$	12.54	0.0003	0.121
	X	7.71	0.0054	0.189
Low Racial Distance $\|\Delta$ Sh. White$\| < 3.1$ p.p.	$Z_{3,gt}$	0.66	0.4144	0.542
	$Z_{2,gt}$	4.06	0.0438	0.342
	$Z_{1,gt}$	8.36	0.0038	0.237
	X	10.30	0.0013	0.093

Note: In orange, the tests for which the p-value is below 0.05, indicating that we can reject the exogeneity of the tested variable with 95% confidence. The J-statistic is computed from the 2-step GMM estimation of α, using $(Z^*_{gt}, \tilde{Z}^{\pm}_{k,gt})$, $k = 1, 2, 3$ as instruments, and allowing for two-way clustering at the origin and destination microregions. The p-value is the probability that a random variable with a $\chi^2(1)$ distribution takes a value above J. The distance in kilometers refers to the geographical distance between the origin and destination microregions. Internet access is measured as the number of broadband internet connections per capita in the destination microregion $\bar{l}$ in period $t-1$. The racial and educational distances between two locations in a given year is the absolute value of the difference between the two locations' share of white individuals and high school graduates or above, respectively.

The results in the first two panels of Table 1.4 confirm that migrants' information sets about the wage differences between a given destination and their current origin are richer if the destination is geographically closer to the origin. The first panel features very similar results to the ones reported in Table 1.3, namely that we can reject that migrants know any of the four candidate variables when trying to predict the wage difference between a remote microregion and their origin. The second panel, however, shows that once we focus on nearby destinations only, we can no longer reject that migrants know the $t-1$ wage difference at the mesoregion level and the sign of the $t-1$ wage difference at the microregion level.

The remaining panels show similar results along the other selected dimensions, namely that migrants' information set are richer for destinations with a higher level of internet access, a higher population density, lower educational distance and lower racial distance. Note, however, that we reject perfect information for all panels, as indicated by the low p-values obtained when we use X as instrument. In all but one panel, we can also reject that agents know the $t-1$ wage difference at the microregion level ($Z_{1,gt}$). In fact, for the least "favorable" set of destinations according to each of the variables W considered, i.e. the most geographically, educationally or racially distant, least dense or least connected to internet, we can reject that migrants know

any of the candidate variables, reflecting the results in Table 1.2. In particular, the fourth panel shows that we cannot reject that migrants know the $t-1$ wage difference at the microregion level for destinations with a high level of internet access. Given the high predictive power of the $t-1$ wage difference for the current wage difference (the R squared of the first stage is 0.828, as reported in column 2 of Table 1.2), this indicates that migrants' information sets are quite rich when they predict the wage difference about locations with high internet access. This is in contrast to the apparent paucity of migrants' information sets when predicting wage differences with other destinations.

In addition to the simple outcome of the overidentification tests, the overall larger p-values for the "favorable" destinations provide further evidence of the heterogeneous pattern of information sets. In all panels, the p-values tend to be larger for the least geographically, educationally and racially distant, most dense and with better internet access destinations, even when we still reject the overidentification test. For example, the p-value for the $t-1$ wage difference for destinations at low educational distance and the p-value for the sign of the $t-1$ wage difference for destinations at low racial distance are both close to 5 percent.

Heterogeneity Across Groups

In Table 1.5, we test whether workers may have better information about earnings in any destination, depending on which group they belong to. Specifically, we test for heterogeneity along characteristics of the origin location and demographic types.

The first four panels indicate substantial heterogeneity in migrants' information sets depending on whether they are located in a microregion with a high or low access to internet, and a high or low population density. Although we always reject perfect information, locations with the highest level of internet access and higher population density are the ones for which we cannot reject that agents know the $t-1$ wage

differences. Since we are selecting on one of these variable at once, it is very likely that origin locations with the highest internet access are also the ones with the highest population density. In order to further assess the relevant source of variation, we will interact population density and internet access in the next section.

We find overall very limited evidence of heterogeneity in migrants' information sets depending on their demographic types. For all but one of the panels studying heterogeneity along one of the demographic categories, we can reject that migrants know any of the candidate variables. We do obtain some evidence in favor of a slightly richer information set for white individuals in comparison to non whites, although the p-values for the test of $Z_{1,gt}$ (0.0587 and 0.0281, respectively) are quite close despite being on either side of the 5 percent threshold. After further inspecting the level of the p-values, the test of $Z_{1,gt}$ is only barely rejected for high school graduates (the p-value is 0.049), while it is frankly rejected for less than high school graduates. This would point toward a slightly richer information set for high school graduates.

By looking at broad demographic categories (male versus female, white versus non white, etc.), each of the tests discussed above imposes the assumption of knowledge of a variable for all individuals of this category. One potential reason why we do not find evidence of heterogeneity of information sets by demographic type may be due to the fact that this heterogeneity is present at a finer level. For example, if only white males with a high school or above educational level have a richer information set, we should expect any of the tests described in Table 1.5 to be rejected. To investigate this further, in the following section we perform tests by interacting some of the categories defined in Table 1.4 and Table 1.5.

Interactions

We now turn to the analysis of the heterogeneity of information sets according to categories that are finer than the ones studied in sections 1.6.2 and 1.6.2. In particular,

Table 1.5: Testing Content of Information Set - Heterogeneity Across Groups

Regions Tested	Variable Tested	J-test Statistic	J-test p-value	$\hat{\alpha}$
Low internet access at origin Sh. Internet < 6.3 %	$Z_{3,gt}$	4.10	0.0428	0.491
	$Z_{2,gt}$	15.11	0.0001	0.159
	$Z_{1,gt}$	13.91	0.0001	0.089
	X	10.25	0.0013	0.181
High internet access at origin Sh. Internet > 16 %	$Z_{3,gt}$	1.10	0.2924	0.445
	$Z_{2,gt}$	0.35	0.5527	0.547
	$Z_{1,gt}$	0.24	0.6209	0.479
	X	6.46	0.0110	0.123
Low population density at origin Pop. dens < 11 inh/km^2	$Z_{3,gt}$	10.44	0.0012	0.460
	$Z_{2,gt}$	17.67	2.6e-05	0.079
	$Z_{1,gt}$	16.05	6.1e-05	0.007
	X	12.02	0.0005	0.114
High population density at origin Pop. dens > 195 inh/km^2	$Z_{3,gt}$	0.83	0.3617	0.566
	$Z_{2,gt}$	0.85	0.3555	0.445
	$Z_{1,gt}$	1.70	0.1917	0.408
	X	4.31	0.0377	0.233
Less than high school graduate	$Z_{3,gt}$	7.37	0.0066	0.494
	$Z_{2,gt}$	6.54	0.0105	0.318
	$Z_{1,gt}$	11.42	0.0007	0.192
	X	8.78	0.0030	0.178
Above high school graduate	$Z_{3,gt}$	4.10	0.0429	0.534
	$Z_{2,gt}$	6.68	0.0097	0.257

	$Z_{1,gt}$	10.67	0.0010	0.133
	X	8.61	0.0033	0.070
Male	$Z_{3,gt}$	4.84	0.0278	0.510
	$Z_{2,gt}$	7.21	0.0072	0.254
	$Z_{1,gt}$	10.43	0.0012	0.143
	X	8.63	0.0033	0.109
Female	$Z_{3,gt}$	5.93	0.0149	0.472
	$Z_{2,gt}$	4.45	0.0348	0.368
	$Z_{1,gt}$	11.33	0.0007	0.214
	X	8.39	0.0037	0.114
White	$Z_{3,gt}$	3.57	0.0587	0.535
	$Z_{2,gt}$	5.91	0.0149	0.274
	$Z_{1,gt}$	9.79	0.0017	0.155
	X	8.83	0.0029	0.063
Non White	$Z_{3,gt}$	4.82	0.0281	0.490
	$Z_{2,gt}$	7.88	0.0049	0.356
	$Z_{1,gt}$	12.62	0.0003	0.207
	X	7.85	0.0050	0.216

Note: In orange, the tests for which the p-value is below 0.05, indicating that we can reject the exogeneity of the tested variable with 95% confidence. The variable $dist_{\bar{l}l}$ denotes the distance (in kilometers) between microregions $\bar{l}$ and l. The variable $internet_{\bar{l}t}$ denotes the number of broadband internet connections per capita in microregion $\bar{l}$ in period $t-1$.

we present the results from interacting internet access at the origin with population density at the origin, and interacting demographic types. Many other combinations

of demographic types, origin and destination characteristics could be studied and we present more such exercises in Appendix A.3.3.

Given the stark heterogeneity in the composition of migrants' information sets depending on the internet access at the origin level and on population density at the origin, it seems natural to ask whether population density, internet access, or both, are driving this heterogeneity. To explore these interactions, we define the following four instruments

$$\tilde{Z}^{++}_{k,glt} \equiv \mathbb{1}\{W_{1,glt} > q^{W_1}_{75}, W_{2,glt} > q^{W_2}_{75}\} Z_{k,glt} + \mathbb{1}\{W_{1,glt} \leq q^{W_1}_{75}, W_{2,glt} \leq q^{W_2}_{75}\} Z^*_{glt},$$

$$\tilde{Z}^{+-}_{k,glt} \equiv \mathbb{1}\{W_{1,glt} > q^{W_1}_{75}, W_{2,glt} < q^{W_2}_{25}\} Z_{k,glt} + \mathbb{1}\{W_{1,glt} \leq q^{W_1}_{75}, W_{2,glt} \geq q^{W_2}_{25}\} Z^*_{glt},$$

$$\tilde{Z}^{-+}_{k,glt} \equiv \mathbb{1}\{W_{1,glt} < q^{W_1}_{25}, W_{2,glt} > q^{W_2}_{75}\} Z_{k,glt} + \mathbb{1}\{W_{1,glt} \geq q^{W_1}_{25}, W_{2,glt} \leq q^{W_2}_{75}\} Z^*_{glt},$$

$$\tilde{Z}^{--}_{k,glt} \equiv \mathbb{1}\{W_{1,glt} < q^{W_1}_{25}, W_{2,glt} < q^{W_2}_{25}\} Z_{k,glt} + \mathbb{1}\{W_{1,glt} \geq q^{W_1}_{25}, W_{2,glt} \geq q^{W_2}_{25}\} Z^*_{glt},$$

where $(W_1, W_2) = \{internet_{\bar{l}t}, popdensity_{\bar{l}t}\}$. That is, we impose the assumption that any given candidate variable belongs to migrants' information sets alternatively in locations with high internet access and high population density, high internet access and low population density, low internet access and high population density, or low internet access and low population density.

The first four panels in Table 1.6 show the results of testing for the validity of each candidate instrumental variable for the four categories of origin locations. The first panel shows that in regions with both low population density and low internet access, we can reject that any of the variables is included in migrants' information sets. The second panel shows that this is also the case in locations that have high population density but low access to internet. Hence, being located in a microregion with high population density does not appear to be sufficient for the information set

to contain any of the candidate variables. In the third panel, we see that in locations with high internet access but low population density, we cannot reject that agents have the knowledge of $Z_{3,gt}$, $Z_{2,gt}$ and even $Z_{1,gt}$. This is a further indication that internet access seems to be the most important driver of heterogeneity in information sets. Reassuringly, we also cannot reject that information sets contain $Z_{3,gt}$, $Z_{2,gt}$ and $Z_{1,gt}$ in microregions that have both a high population density and high internet access. Note that even in these most "favorable" locations, we can still reject perfect information with a pretty high level of confidence (the p-value is 0.0049). These four panels show that the role of internet access at the origin level seems to operate even when once we control for population density, and that the heterogeneity along population density described in Table 1.5 may well be driven by the positive correlation between internet access and population density.

The next four panels in Table 1.6 show the results of studying the heterogeneity of information sets by interacted education and race categories. This exercise is motivated by the observation in Table 1.5 that educational attainment and race were the two demographic characteristics along which some (mild) heterogeneity in information sets appeared. The results for non white individuals with less than high school education level show that we can reject any of the candidate variables. This is also true for non white with high school or above education level. However, the p-value for the level of the wage difference at the mesoregion level ($Z_{3,gt}$) is close to the threshold (it is equal to 0.0450). As in Table 1.5, this points towards some mild heterogeneity in information sets along educational attainment for non white individuals. The second-to-last panel shows that for white individuals with less than high school education level, we cannot reject that the level of the wage difference at the mesoregion level ($Z_{3,gt}$) belongs to their information set. Finally, the last panel confirms that individuals with higher education attainment have richer information sets, since we cannot reject that white individuals who are high school graduates or above know

Table 1.6: Testing Content of Information Set - Interactions

Regions Tested	Variable Tested	J-test Statistic	J-test p-value	$\hat{\alpha}$
Low pop. dens. & Low internet at orig.	$Z_{3,gt}$	6.44	0.0112	0.496
	$Z_{2,gt}$	7.02	0.0080	0.254
	$Z_{1,gt}$	10.84	0.0009	0.132
	X	8.64	0.0032	0.086
High pop. dens. & Low internet at orig.	$Z_{3,gt}$	7.91	0.0049	0.507
	$Z_{2,gt}$	10.83	0.0009	0.161
	$Z_{1,gt}$	12.28	0.0004	0.078
	X	9.91	0.0016	0.101
Low pop. dens. & High internet at orig.	$Z_{3,gt}$	0.34	0.5562	0.513
	$Z_{2,gt}$	2.76	0.0966	0.479
	$Z_{1,gt}$	2.98	0.0843	0.473
	X	5.46	0.0195	0.421
High pop. dens. & High internet at orig.	$Z_{3,gt}$	0.29	0.5869	0.484
	$Z_{2,gt}$	3.08	0.0789	0.381
	$Z_{1,gt}$	3.33	0.0680	0.422
	X	7.90	0.0049	0.103
Non White Less Than High School	$Z_{3,gt}$	7.01	0.0081	0.415
	$Z_{2,gt}$	6.21	0.0126	0.386
	$Z_{1,gt}$	11.15	0.0008	0.307
	X	7.61	0.0057	0.281
Non White Above High School	$Z_{3,gt}$	4.02	0.0450	0.507
	$Z_{2,gt}$	8.48	0.0035	0.362

	$Z_{1,gt}$	12.99	0.0003	0.211
	X	7.88	0.0049	0.220
White Less Than High School	$Z_{3,gt}$	0.26	0.6084	0.530
	$Z_{2,gt}$	5.45	0.0194	0.335
	$Z_{1,gt}$	9.86	0.0016	0.225
	X	8.94	0.0027	0.179
White Above High School	$Z_{3,gt}$	3.32	0.0682	0.493
	$Z_{2,gt}$	3.28	0.0701	0.441
	$Z_{1,gt}$	9.57	0.0019	0.162
	X	8.92	0.0028	0.045

Note: In orange, the tests for which the p-value is below 0.05, indicating that we can reject the exogeneity of the tested variable with 95% confidence. The internet measure is the number of broadband internet connections per capita at $t-1$.

$Z_{3,gt}$ and $Z_{2,gt}$. These four panels illustrate that individuals that are white appear to have slightly richer information sets than non white individuals, and that individuals that are high school graduates or above also have slightly richer information than less educated individuals.

1.7 Concluding Remarks

In this paper, we study the extent to which workers' limited migration responses to local labor demand shocks are due to their lack of information about their potential net gains from regional migration. Using data for the period 2000 to 2014 on the location of work of every legally employed Brazilian worker, we first present suggestive evidence in support of information frictions affecting migration decisions. We show that in response to several positive local labor demand shocks that occurred in Brazil

between 2006 and 2010, migrants coming from regions that are located farther away from the shock and with lower access to internet tend to move later, even controlling for several characteristics of individuals and of their region of origin.

In the second part of the paper, we propose an empirical approach to estimate information frictions in a context where migration decisions result from forward-looking agents maximizing their expected utility from moving to any location. Using our revealed preference method, we first show that estimates of the elasticity of migration with respect to wages, a parameter with key welfare implications in structural analyses of migration patterns, are highly dependent on the assumptions about the information available to migrants. Our preferred estimates, that rely only on agents knowing a the sign of the wage difference at the mesoregion level in the previous year, are almost three times larger than the estimates obtained under the common assumption that agents know the exact wage difference at the microregion level in the previous year.

The results from the implementation of our testing procedure reveal two key patterns. First, in stark contrast with the common assumption in the literature of perfect information about wages, we find overwhelming evidence of incomplete information, for all groups of workers and all possible destinations. Second, workers' information sets are heterogeneous, with individual demographics, characteristics of the place in which they reside and of the place they are considering affecting the extent to which they are informed about the annual wages. Our results show that gravity in terms of geographic, educational and racial distance, as well as the degree of internet penetration and population density in the region of residence of a worker crucially determine the extent to which such worker is informed about the net gains of migrating to any other Brazilian region. We also find that workers with a higher educational attainment and who identify as white have richer information sets.

The heterogeneity in workers' information sets that we document is likely to have important welfare implications. Workers living in the best informed regions and part of the best informed demographic categories may have a significant advantage when making their migration decisions. By exploiting their more precise information when choosing where to move (and whether to move), more informed workers are likely to enjoy higher wages and amenity levels on average than less informed workers. In future work, we intend to quantify the welfare consequences of such informational advantages by performing partial equilibrium counterfactual exercises in which we study the response to local shocks of individuals with information sets that are heterogeneous along the dimensions identified in this work. We also intend to extend our exercise to allow for individual-specific information sets and rely on moment inequalities to partially identify the structural parameters of the model and the content of information sets.

Chapter 2

Migration with Costly Information

2.1 Introduction

Migration decisions affect many aspects of workers' lives, from employment opportunities and housing conditions, to schooling and entertainment options. To evaluate migration opportunities, workers must therefore gather information about a wide array of location attributes. Collecting and processing this information entails costs, yet we know little about the nature and magnitude of these costs, and even less about their role in shaping migration decisions. These information costs might depend on where workers live. Workers might share information about past migrations with their neighbors, and some locations may offer better infrastructure for collecting information. Since migration contributes to local networks and relocates workers, the structure of information itself may be altered by migration patterns. How important are information frictions in shaping migration patterns, and conversely, how does migration modify these frictions? What is the scope for policies to affect the structure of information and to induce migrants to take better advantage of migration opportunities?

In this paper, I propose answers to these questions with four contributions. First, I develop a theory of spatial equilibrium with migration in the presence of information frictions, in which the structure of workers' information is determined as an equilibrium outcome. Second, I apply my model to internal migration in Brazil, and structurally estimate it to quantify the magnitude of information costs. Third, I show that my model reproduces the observed heterogeneity of migration elasticities and delay in response to local shocks. Fourth, I quantify the local and aggregate welfare gains from the roll-out of internet in Brazil.

In the first part of the paper, I propose a quantitative dynamic model of migration with costly information acquisition and local information sharing. Each period, unobserved location-specific productivity shocks alter the spatial distribution of earnings. Agents are rationally inattentive and can acquire information at some cost, which can vary by location. Agents use their beliefs to make location decisions, facing fixed bilateral costs of moving between any two locations. Each agent shares their individual information with all agents in the same location. In a given location, agents are heterogeneous along two dimensions. First, they may have different beliefs about the distribution of payoffs in each location. Second, they may have different preferences for each location. While preference-based migration reflects utility maximization, migration decisions under incomplete information are prone to mistakes.

My model allows for a tractable solution of the stochastic steady state and yields three results. First, the bilateral migration shares take a closed-form multinomial logit form akin to existing models of migration. In particular, when the cost of information acquisition converges to zero, the model reduces to existing logit models of migration driven by preference heterogeneity. Second, agents optimally choose to be more informed about locations offering high average payoffs, leading to a bias towards nearby locations. By enhancing flows to closer destinations at the expense of more remote ones, these predispositions act as additional endogenous bilateral migration

costs that depend on distance, as in the gravity literature.[1] Third, for a specific distribution of idiosyncratic preferences that delivers a closed-form aggregation of individual decisions, the model disentangles the effects of idiosyncratic preferences and of the lack of information. Although both reduce the responsiveness of migration flows to variations in earnings, information frictions particularly affect the response to unobserved payoffs. Recovering both the elasticity of migration with respect to unobserved productivity shocks and the elasticity with respect to wages allows me to separately identify information frictions from preference heterogeneity.

Although the model is analytically tractable and delivers closed-form expressions, migration decisions depend on a large number of states. The state variables consist of the vectors of productivity and population in each location, as well as the local belief distributions inherited from previous periods. A complete representation of the beliefs would require to treat the moments of each local belief distribution as state variables. To overcome this challenge, I first show that in the stochastic steady state the beliefs can be described as a function of population only, drastically reducing the effective state space to the productivity and population vectors. Second, since the state space is still too large to employ standard techniques of dynamic programming, I resort to approximate dynamic programming methods (Powell, 2011). Specifically, I use a polynomial approximation for the value function, approximate equilibrium beliefs by the conjugate of a Type 1 extreme value distribution, and use a sample of states in the solution algorithm.[2] I confirm that the algorithm is accurate by comparing it to an almost-exact solution when the number of locations is small or when the productivity process is discrete.

[1] See Monte et al. (2018), Morten and Oliveira (2018), Tombe and Zhu (2019), Fan (2019), Caliendo et al. (2019) for analyses of spatial equilibria featuring preference-based migration or commuting with costly bilateral migration costs, leading to a gravity expression for mobility flows.

[2] See Brown and Jeon (2019) for a similar assumption on the distribution of beliefs, and Nadarajah (2008) and Marques et al. (2015) for a description of the conjugate of the Type 1 extreme value distribution.

In the second part of the paper, I apply my model to internal migration in Brazil and assess the quantitative importance of information frictions. The closed-form expression for migration probabilities allows for a transparent estimation strategy. To estimate the relevant parameters of the model, I rely on detailed migration flows between the 137 Brazilian regions over 15 years. I split these 15 years into two periods, the first from 2000 to 2007; the second from 2008 to 2014. I assume that the economy is in a separate steady-state over each of these periods. I construct migration flows from administrative matched employer-employee data covering the universe of workers employed in the formal sector. I observe workers' location and earnings each year, representing 24 million distinct employees per year on average. I exploit the gravity structure of migration flows predicted by my model to derive regression equations that identify the information costs, preference heterogeneity and migration costs separately. According to the model, migration flows respond differently to changes in payoffs depending on whether agents directly observe these payoffs or if they must acquire information about them. I assume that agents observe population in each region, but do not directly observe local productivity shocks. After recovering the local productivity shocks, I compare the migration elasticity to wages and to local productivity shocks to identify the role of preference heterogeneity and information costs.

The estimated costs of information amount on average to workers paying 3% of earnings each year to acquire information about other regions. The estimated information costs are lower in regions with a higher fraction of residents with an internet connection. Using a region's proximity to backbone internet cables to instrument for the share of residents with an internet connection, I find that increasing the share of residents with an internet connection by 1% reduces the local cost of information acquisition by 0.83%. The annualized average bilateral migration cost is also 3% of earnings, so that deciding where to go appears as costly as moving. Importantly,

the estimated migration costs are 40% smaller than in the case without information costs.[3] In the presence of information frictions, moderate migration costs can rationalize the low observed migration flows: when migration costs to a region are large, agents acquire little information about this region, further reducing their likelihood of moving to the region.

In the third part of the paper, I show that my estimated model successfully predicts two key features of observed migration patterns that the model with complete information cannot. First, allowing for flexible bilateral migration costs, the elasticity of migration is larger for origin-destination pairs that are closer geographically, that have more interactions in the form of larger past migration flows, and for origins with higher access to internet. This higher responsiveness of mobility decisions arises because agents' beliefs are more tightly correlated with the true payoffs in regions with which they have more interactions. Second, in response to a positive local shock in a region, the migration response to this region is slower for origins that are farther away or have lower internet access. In the model, this delay is due to the gradual updating of beliefs about the shock, through local information sharing. In regions that are farther away or that have higher information costs, this local information sharing is slower.

The final part of the paper illustrates the quantitative implications by undertaking two counterfactual exercises. First, I evaluate the effect of removing all information frictions by comparing the steady state equilibrium with the estimated information costs to the case where these costs are set to zero. I find that welfare increases by 5.5%, with a 15% decrease in the cross-sectional dispersion in earnings, reflecting a better arbitrage of local shocks. In the steady state with complete information, gross migration flows are more concentrated towards high payoffs regions and are

[3]The idea that part of the large estimated migration costs could be attributed to information frictions can be traced back at least to Sjaastad (1962). He mentioned: "One is strongly tempted to appeal to market imperfections such as the lack of information to explain the apparently high distance cost of migration."

4.1% lower. Second, I evaluate the counterfactual effect of the roll-out of broadband internet in Brazil during the early 2010s.[4] I compare the outcomes in a steady-state economy where internet is never introduced to an economy in which the change in information costs reflects the average internet access observed between 2008 and 2014. I find that the expansion of internet access increases average welfare by 1.6%. I then decompose these welfare changes into several components. Some gains are mechanically the result of the lower information costs. Second, local networks transmit more information so workers can spend less on individual information acquisition. A third potential source of gains is the better sorting of agents to regions offering higher payoffs. Finally, some gains may arise from a change in regional outcomes, as wages adjust in response to the different spatial allocation of workers. The positive average welfare gains mask substantial heterogeneity in welfare effects, with several remote regions experiencing small welfare losses. Fewer well-informed workers move to these relatively unattractive regions in a steady state with better information access, making local information sharing less effective, and ultimately worsening workers' spatial sorting.

This paper is related to several existing literatures. A recent set of empirical studies emphasizes the importance of information frictions for migration decisions. First, there is growing evidence that migration decisions can be affected by the provision of information. This has been shown by exploiting variation in migrants' access to information about migration opportunities arising either from differential media exposure (McCauley, 2019; Farré and Fasani, 2014; Wilson, 2018), or by directly providing information about average earnings in other locations in a randomized experiment (Baseler, 2019; Bryan et al., 2014).[5] Second, a recent body of work documents that

[4] I abstract from the effects that broadband internet access may have had on local productivity and focus only on its role in improving migrants' information. Of course, there is still an effect of migration on aggregate productivity.

[5] Notably, Bryan et al. (2014) found that in the context of seasonal migration in Bangladesh, providing information about average wages and availability of jobs in four broad regions did not result in any significant increase in migration. They concluded that either households already had

migrants tend to have inaccurate information about the returns to migration. In the context of international migration, a number of papers directly measured migrants' expectations in surveys, lab, and randomized field experiments, and find that, for most migrants, they were largely misaligned with actual outcomes (McKenzie et al., 2013; Bah and Batista, 2018; Shrestha, 2017). In an analysis of internal migration in Brazil, Fujiwara, Morales and Porcher (2020) use a revealed preference approach to infer the composition of migrants' information consistent with a model of migration with a rich structure of migration costs. They find that the structure of information needed to rationalize the observed migration flows is concentrated on a few neighboring regions and larger cities, with poor information about regions beyond several hundred kilometers. My paper contributes to this literature by providing a general equilibrium theory of optimal information acquisition that leads to an endogenous structure of information. This information structure is also concentrated on nearby regions and larger cities. This framework allows me to quantify the implications of these information frictions for aggregate welfare.

This paper also contributes to a rapidly growing empirical literature on information transmission through social networks (Granovetter, 1973). A recent series of papers shows that workers use information obtained from their coworkers (Dustmann et al., 2015; Glitz and Vejlin, 2019; Saygin and Weynandt, 2014; Caldwell and Harmon, 2019), family members (Kramarz and Skans, 2014), neighbors (Bayer et al., 2008; Schmutte, 2015) and classmates (Zimmerman, 2019) to find job opportunities. This paper provides suggestive evidence that workers also rely on their social networks to gather information relevant for migration decisions.

This paper is related to the emerging theoretical literature on rational inattention in the context of discrete choice. Following the seminal contribution by Sims (2003),

this information, or the information made available was not useful or credible. Consistent with the first interpretation, Fujiwara, Morales and Porcher (2020) cannot reject that internal migrants in Brazil have knowledge of average wages at a broad regional level. However, they seem to lack information about labor market outcomes at a finer geographical level.

Matejka and McKay (2015) showed how static rational inattention problems lead to a multinomial logit decision rule.[6] Steiner et al. (2017) extended this property to single-agent dynamic problems, opening a pathway towards incorporating rational inattention in richer dynamic settings. I combine their results with properties of social learning in networks derived by Molavi et al. (2018) to show that the dynamic logit structure survives in steady-state in environments with a continuum of agents, heterogeneous preferences, and endogenous payoffs. There is so far very limited work incorporating the rational inattention framework into structural models. Some exceptions in the industrial organization literature include Joo (2017) and Brown and Jeon (2019), while in international trade Dasgupta and Mondria (2014) provided a microfoundation for the model from Eaton and Kortum (2002).[7] In the first application to migration, Bertoli et al. (2019) show that international migration flows feature heterogeneous elasticities, consistent with rational inattention.[8] I provide the first structural estimation of a dynamic rational inattention model.

My paper also contributes to the quantitative economic geography literature. A number of studies have documented low migration responses in reaction to local shocks.[9] In their analysis of internal migration patterns in the United States, Ken-

[6]Other important contributions to the analysis of rational inattention problems in discrete choice include Caplin and Dean (2015); Caplin et al. (2019). Fosgerau et al. (2019) demonstrate a general equivalence between the class of additive random utility models and rational inattention problems with generalized entropy.

[7]Brown and Jeon (2019) are the first to offer a tractable combination of preference heterogeneity with rational inattention in a static framework by assuming that the beliefs and idiosyncratic preference shocks are described by the conjugate of a Type 1 extreme value distribution.

[8]They show, in particular, that migration flows emanating from origin countries with highly concentrated flows to one destination, such as Mexico to the United States, are less responsive to wages in other destinations. They discuss how this lower migration elasticity can be rationalized by a lower value of full information for migrants from Mexico than for migrants from countries with many potential destinations, such as China.

[9]Following early studies of local labor demand shocks (Bartik, 1991; Blanchard and Katz, 1992), limited migration responses have been observed in reaction to international trade shocks (Topalova, 2010; Kovak, 2013; Autor et al., 2013; Adão, 2016; Dix-Carneiro and Kovak, 2017; Pierce and Schott, 2018; Dix-Carneiro and Kovak, 2019; Adão et al., 2019), changes in technology (Bustos et al., 2016; Acemoglu and Restrepo, 2017), local shocks to housing net worth (Mian and Sufi, 2014), differential incidence of business cycles (Yagan, 2019; Beraja et al., 2019), place-based policies (Busso et al., 2013), natural resource discovery (Bartik et al., 2019) and natural disasters (Nakamura et al., 2019).

nan and Walker (2011) point out that large migration costs are a priori necessary to explain the concurrence of important spatial disparities in incomes, a sizeable elasticity of migration with respect to income variations, and overall limited migration flows.[10] Such large migration costs have been shown to have important implications for aggregate labor productivity and welfare.[11] My analysis illustrates how information frictions affect the migration elasticity as well as migration costs and suggests that as much as 40% of the migration costs estimated under the assumption of complete information could in fact be attributed to information frictions. Recent analyses have emphasized the relevance of spatial linkages due to trade and labor mobility for the adjustment of economies to various shocks (Monte et al., 2018; Tombe and Zhu, 2019; Caliendo et al., 2019; Adão et al., 2019). By incorporating the role of information frictions in a dynamic spatial equilibrium model, this paper describes how the endogenous structure of information interacts with the spatial allocation of economic activity, creating an additional channel of adjustment with important welfare implications.

In Section 2.2, I present my model. Section 2.3 describes the data I will be using for the estimation. In Section 2.4, I carry out the estimation and discuss the results. Section 2.5 confronts some of the model's predictions to the data. Section 2.6 presents the counterfactual exercises.

[10] Kennan and Walker (2011) estimate that moving costs above $300 thousand 2010 dollars on average are needed to account for observed migration flows across U.S. states. Diamond et al. (2019) find fixed cost of moving between neighborhoods in San Francisco of around $40 thousand.

[11] For quantifications of the aggregate implications of workers' limited geographic mobility, see Redding (2016); Diamond (2016a); Morten and Oliveira (2018); Bryan and Morten (2018); Caliendo et al. (2018, 2019).

2.2 A Model of Migration with Costly Information Acquisition and Local Information Sharing

In this section, I present the dynamic general equilibrium model of migration with costly information acquisition and local information sharing. After describing the environment and the structure of flow payoffs, I present the three main blocks of the model, consisting of the individual information acquisition problem faced by rationally inattentive agents, their mobility decision, and local information sharing. Combining these elements, I characterize the steady-state equilibrium and discuss the approximated dynamic methods used for the simulation.

2.2.1 Set Up

The objective of the model is to capture how agents make location choices in an environment in which at least some component of local payoffs varies over time, and in which it may be too difficult for agents to track these fluctuations perfectly. Therefore, agents face a trade-off between the value of holding precise information about current payoffs and the costs of gathering such information.

With this goal in mind, I consider an infinite-horizon environment with J locations. Each location j is characterized by a time-invariant productivity level A_j and amenities B_j. The cross-sectional variation in baseline productivities A_j helps explain persistent spatial dispersion in earnings, and the variation in amenities B_j justifies why some locations attract more workers than other locations with similar earnings levels. Each period, locations experience an exogenous stochastic productivity shock θ_{jt}, meant to capture the fluctuations in migration opportunities over time.[12] Let

[12] As discussed in Footnote 9, such fluctuations in local labor demand may be due to locations' differential exposure to international trade competition, changes in technology, government spending, business cycles, the location of firms, or natural resource discovery that I do not model. In Section 2.5, I focus on a number of such local shocks in Brazil between 2000 and 2014, including dam construction, mining and oil booms, and surges in tourism activity.

$\Gamma(\theta_t|\theta^{t-1})$ denote the distribution of the vector $\theta_t = (\theta_{1t}, \dots, \theta_{Jt})$, which can depend on past realizations $\theta^{t-1} = (\theta_{t-1}, \theta_{t-2}, \dots)$. The geography is represented by a set of fixed bilateral migration costs κ_{jk} that agents must pay if they decide to move from any location j to another location k. These costs are common to all agents in the same location, reflecting both the fiscal cost of moving as well as utility costs, such as being away from friends and family (Sjaastad, 1962).

Agents who start period t in some location j choose where to locate for the rest of the period. Before moving, they draw a vector of idiosyncratic preferences for each location. Denoting by $u(c_{kt})$ the flow of utility from consumption derived by agents in location k in period t, I can then write the gross flow of utility u_{ijkt} for an agent i moving from j to k at t as

$$u_{ijkt} = u(c_{kt}) + B_k - \kappa_{jk} + \nu\varepsilon_{ikt}, \tag{2.1}$$

reflecting the utility gain from consumption and amenities in the destination minus the mobility cost, where ε_{ikt} represents agent i's idiosyncratic taste for location k at t, scaled by the parameter ν. I assume that ε_{ikt} are identically and independently distributed across individuals i, locations k and periods t.

In order to make the role of information frictions more salient in the exposition of the model below, I maintain simple assumptions on the production side of the economy. There is a unique freely traded homogeneous good chosen as numeraire. I assume that the production function takes a Cobb-Douglas form using labor as the single input, so that output in location k at t is given by

$$y_{kt} = \exp(A_k + \theta_{kt}) L_{kt}^{1-\alpha}, \qquad \alpha > 0, \tag{2.2}$$

where θ_{kt} is the current productivity shock in location k, and L_{kt} is the human capital in location k at time t, equal to the population in k since I assume every agent supplies

one unit of human capital. The parameter α captures the decreasing marginal product of labor. I assume that the vector θ_t follows an AR(1) process with persistence ρ and variance of the innovation σ_ξ^2, so that in each period t the productivity shock in location j is related to the previous period shock according to

$$\theta_{jt} = \rho\theta_{jt-1} + \xi_{jt}, \qquad \xi_{jt} \sim \mathcal{N}\left(0, \sigma_\xi^2\right). \tag{2.3}$$

The parameter α, governing decreasing returns in labor in production, introduces congestion to the model, which will reduce the attractiveness of a location as more workers move in. More generally, congestion could also arise because of a fixed or imperfectly elastic supply of housing or land. A positive value for α implies that the level of earnings in each location depends on, and affects, the population level resulting from migration choices. The first order condition arising from profit maximization yields the following expression for the wage in location k at t:

$$\log w_{kt} = A_k + \theta_{kt} - \alpha \log L_{kt} + \log\left(1-\alpha\right). \tag{2.4}$$

In this economy, firms make profits. I assume that profits made in each location are collected and redistributed to workers via a negative tax rate on their local wage τ_t, constant across locations, so that earnings in location k at time t are $w_{kt}\left(1+\tau_t\right)$. The log-linear structure of utility implies that rebating profits does not alter migration decisions, since earnings increase by the same rate in every location. It easy to show that the rate of transfer is constant over time and equals $\tau = \alpha/(1-\alpha)$, so the net-of-transfer earnings in location k are $w_{kt}/(1-\alpha)$.[13] Moreover, the assumption that the economy features free trade in goods between locations implies that local consumption prices are equalized across locations and do not affect migration decisions. As a result,

[13]The profits made in location k are equal to $\alpha \exp\left(A_k + \theta_{kt}\right) L_{kt}^{1-\alpha}$. The optimal transfer rate τ_t then solves $\sum_k \tau_t w_{kt} L_{kt} = \sum_k \alpha \exp\left(A_k + \theta_{kt}\right) L_{kt}^{1-\alpha}$, leading to $\tau = \alpha/(1-\alpha)$ after substituting wages by their expression in (2.4).

indirect utility net of information costs can be expressed as

$$u_{ijkt} = A_k + \theta_{kt} - \alpha \log L_{kt} + B_k - \kappa_{jk} + \nu\varepsilon_{ikt}. \tag{2.5}$$

2.2.2 Timeline of the Model

I now describe the sequence of actions taken by agents in each period. I assume that the stochastic productivity vector θ_t is the only imperfectly observed variable.[14] Although the population distribution is a time-varying object, it is arguably easier for agents to have knowledge of the population rather than the productivity shock in each location. It may also seem plausible that local amenities are also subject to stochastic variation. Given that local productivity shocks are the only variables not directly observed, agents hold beliefs about the cross-sectional distribution of productivity. Denote by π_{it} the prior beliefs about the vector θ_t held by agent i at the beginning of period t:

$$\pi_{it}(\theta) = \Pr(\theta_t = \theta | i, t), \qquad \forall \theta \in \mathbb{R}^J. \tag{2.6}$$

The assumption that productivity shocks follow an AR(1) process implies that the dependence of the current shock on past realization can be summarized by the shock in the previous period. Hence, agents only use their beliefs about the previous period productivity to form beliefs about the current distribution of shocks.

In this economy, agents' decisions depend on several variables. First, agents with different prior beliefs π_{it} may decide to acquire different information and make different mobility decisions. Second, the distribution of payoffs available to agents depends on the population inherited from the previous period L_{t-1}, the location $l_{it-1} \in \{1, \ldots, J\}$ in which they start the period—which determines the migration

[14] One justification for this is that time-invariant variables such as amenities, and baseline productivities amenities have been learned gradually over time until they became perfect knowledge.

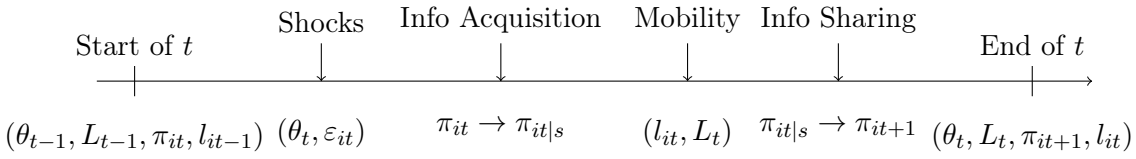


Figure 2.1: Sequence of actions during any period t

costs they face—and their current preference shocks ε_{it}. I collect these state variables into $\omega_{it} = (L_{t-1}, \pi_{it}, l_{it-1}, \varepsilon_{it})$. Finally, agents' decisions also depend on the realization of productivity θ_t, since it will affect the signals that they receive.

As represented in Figure 2.1, each period has four steps. First, the productivity vector θ_t and idiosyncratic preferences ε_{it} are realized. Second, agents can refine their prior beliefs by acquiring and processing information. This step is governed by the rational inattention problem described in Section 2.2.3, and leads agents to form posterior beliefs about θ_t, denoted by $\pi_{it|s}$, where s represents the signal received by agent i. Third, agents use their posterior beliefs to compute the expected payoffs in every location, and move to the location offering the highest expected payoff. The new location l_{it} realized for each agent leads to a new distribution of the population L_t, described in Section 2.2.4. Once agents reach their destination, they engage in local information sharing. Under simple assumptions, I show that all agents in a location reach a consensus about the distribution of θ_t after they communicate their beliefs to each other. This location-specific belief can be used to form beliefs π_{it+1} about θ_{t+1}.

2.2.3 Individual Information Acquisition

I now describe the rational inattention problem. Agents face a trade off between making accurate predictions about the payoffs in each location and paying the cost of acquiring and processing data to make such predictions.

Following Matejka and McKay (2015) and Steiner et al. (2017), individual i in location j starts with a prior π_{it} about θ_t and rationally chooses how much information to acquire about θ_t to form posterior beliefs about the payoffs in each location. Agent i can choose to receive signals s about the current θ_t, which will allow her to form more precise posterior beliefs. Agents cannot control the realization of the signals they receive, but are free to choose the distribution from which they are drawn. Given their prior π_{it} and other observed variables that form their observed state ω_{it}, agents choose the conditional distribution $f(s|\theta_t, \omega_{it})$. Their information acquisition strategies $f(\cdot)$ are unconstrained, reflecting the idea that agents can gather information in many different ways. Although agents are free to design any signal structure, there is a cost of acquiring information so that more "informative" signal structures are more costly. As a result, individuals may wish to become partially informed about locations, i.e. receive a vector of signals with limited information. Once an agent chooses her signal structure for the period, nature draws a signal realization s from $f(\cdot)$. Given the signal, the agent updates her prior, resulting in the posterior belief $\pi_{it|s}$ after applying Bayes' rule:

$$\pi_{it|s}(\theta_t) = \frac{f(s|\theta_t, \omega_{it})\pi_{it}(\theta_t)}{f(s|\omega_{it})}. \tag{2.7}$$

To describe the costs of acquiring information, I follow the rational inattention literature and rely on the entropy of beliefs to measure their uncertainty.[15] For any random variable X with continuous support S distributed according to $p \in \Delta(S)$, the entropy of X, or equivalently the entropy of the distribution p, is defined as

$$H(p) \equiv \tilde{H}(X) = -\int_{x \in S} p(x) \log p(x) dx. \tag{2.8}$$

[15] More specifically, I rely on the Shannon entropy, first introduced in this literature by Sims (2003). More recently, Fosgerau et al. (2019) study information costs based on a more general class of entropy functions.

Entropy is a measure of uncertainty about X.[16] Starting from a prior belief distribution π_{it}, a signal distribution that reduces the expected entropy of θ_t is more costly. To capture this idea, I assume that the cost of a signal distribution f is proportional to the difference between the entropy of the prior beliefs and the expected entropy of the posterior beliefs.[17] I define the cost of designing a signal strategy f for agent with state ω_{it} to be

$$I\left(f|\omega_{it}\right) = \lambda_{l_{it-1}}\left(H\left(\pi_{it}\right) - \mathbb{E}_s\left[H\left(\pi_{it|s}\right)\right]\right), \tag{2.9}$$

so that the cost of a given information strategy f is higher the more it is expected to reduce the agent's uncertainty about θ_t once the signal s is received. Before receiving s, the uncertainty can be measured by $H\left(\pi_{it}\right)$. Afterwards, it becomes $H\left(\pi_{it|s}\right)$. The location-specific parameter $\lambda_{l_{it-1}}$ scales this information cost, reflecting that some locations may offer more efficient technologies to gather information and reduce uncertainty.

An information strategy therefore assigns a signal distribution $f_t(s_t|\omega_{it})$ for each productivity θ_t, population distribution L_{t-1}, prior beliefs π_{it}, location l_{it-1}, and preference draws ε_{it}. Agents also devise mobility strategies $\sigma_t(s_{it}, \omega_{it})$ indicating the choice of location at time t for each current costly signal s_{it}, population distribution L_{t-1}, prior beliefs π_{it}, origin location l_{it-1}, and preference draws ε_{it}, such that agents solve

$$\max_{f,\sigma} \mathbb{E}\left[\sum_{t=1}^{\infty} \delta^t \left(u_{l_{it-1}l_{it}}(\theta_t, L_t, \varepsilon_{it}) - I(f_t|\omega_{it})\right)\right], \tag{2.10}$$

where $l_{it} = \sigma_t(s_{it}, \omega_{it})$ is the location optimally chosen at t, the flow of utility is defined in (2.5), the information cost is defined in (2.9), and the expectation is taken

[16] For instance, the smallest value of entropy of zero is obtained for a Dirac distribution assigning a probability one to some value and zero to all others – with the convention that $0 \log 0 = 0$. For a normal distribution, the entropy is $\log\sqrt{2\pi e\sigma^2}$ and increases with the variance σ^2.

[17] The difference between the entropy of the prior and the expected entropy of the posterior is called the *conditional mutual information* between s_t and θ_t.

with respect to all possible future realizations of productivity and individual states, using current beliefs.

In this economy, agents never see the actual realization of the productivity vector θ_t. In particular, I assume that agents do not use the wage they receive in their current location k to update their local beliefs about θ_{kt}. If the productivity shocks are iid over time, knowing θ_{kt} after moving is not helpful for future migration decisions. However, when the persistence parameter ρ is positive, learning about θ_{kt} allows agents to form more precise beliefs about θ_{kt+1}. In Appendix B.1.2, I provide an extension of my model in which the observation of the wage allows workers to update their beliefs about local productivity. If no additional unobserved shock affects wages, then workers in location k can infer θ_{kt} exactly after they move there. I show, in a simulation with a small number of locations, that for a reasonable set of parameters, the mobility decisions are only mildly affected by the perfect observation of the local θ_{kt}.[18]

2.2.4 Mobility

Once agents have acquired information and formed posterior beliefs $\pi_{it|s}$, they use these beliefs to compute the expected payoff in each location. They then move to the location l_{it} offering the highest expected payoffs. These mobility decisions lead to a new allocation of workers across locations, namely a new population distribution L_t. Denote by $p_{jkt}(s, \theta_t, \omega_{it})$ the probability that an agent i in location $l_{it-1} = j$ would move to location k at time t if the realized productivity is θ_t and the signal is $s_{it} = s$. Using the optimal information f_t and mobility strategy σ_t, this probability is

[18] This insensitivity is due to the endogenous adjustment of agents' information strategies. Since agents are able to form more precise prior beliefs about θ_{kt}, they choose to invest less attention about θ_{kt}. This limits the advantage provided by the costless observation of wages by making mobility less to the realization of θ_{kt}. Even though their choices are moderately affected, agents are still better off, since they save on information acquisition. The crowding out of costless signals on information acquisition was first pointed out by Steiner et al. (2017).

$$p_{jkt}(s, \theta_t, \omega_{it}) = f_t(s|\theta_t, \omega_{it}) \mathbb{1}\{\sigma_t(s, \omega_{it}) = k\} \mathbb{1}\{l_{it-1} = j\}. \tag{2.11}$$

I assume that the total population in the economy stays constant at $\bar{L}$, so that the population distribution evolves in the set $\mathcal{L} = \{\{L_{jkt}\}_{j,k} | \sum_{j,k} L_{jkt} = \bar{L}, L_{jkt} \geq 0\}$. The population in any location k after migration is then $L_{kt} = \sum_j L_{jkt}$, where

$$L_{jkt} = L_{jt-1} \bar{p}_{jkt}(\theta_t, L_{t-1}), \tag{2.12}$$

and $\bar{p}_{jkt}(\theta_t, L_{t-1}) = \mathbb{E}\left[p_{jkt}(s, \theta_t, \omega_{it})\right]$ is the expected probability of moving from j to k at time t, and the expectation is taken over all possible beliefs π_{it}, signals s, and preference shocks ε_{it}.

In order to forecast the wages in every location, agents form rational expectations about the behavior of other agents. The wage w_{kt} in each location k depends on the equilibrium population L_{kt} after all agents have made their mobility decisions. Agents' understanding of the economy, including the distribution of preference shocks and productivity shocks, allows them to successfully predict the mobility flows to every location $\bar{p}_{jkt}(\theta_t, L_{t-1})$, conditional on the observed population distribution L_{t-1} and a particular guess for θ_t.[19]

Although in typical datasets we do not observe individual beliefs π_{it}, signals s, and preference shocks ε_{it}, bilateral migration flows between locations are the empirical equivalent of the aggregate migration probability $\bar{p}_{jkt}(\theta_t, L_{t-1})$. My goal is to obtain an expression for $\bar{p}_{jkt}(\theta_t, L_{t-1})$.

[19] In the same way as observing wages w_{kt} at the end of period t could in principle allow workers to refine their beliefs by inverting the wage-setting equation (2.4), the observation of the vector of mobility flows L_t at the end of period t could be used by agents to back out information about the realized vector θ_t by inverting the law of motion for the population (2.12). In Appendix B.1.2, I discuss the implications of allowing agents to exactly recover the realization of the whole vector θ_t upon observing the population distribution L_t. Agents have the same prior beliefs about θ_{t+1} in every location, but still design different information strategies depending on their location and preference shocks, leading to migration decisions that still feature an important role for information frictions.

2.2.5 Local Information Sharing

In practice, workers can potentially benefit from their local interactions to obtain information relevant for future mobility decisions. For example, individuals who decided to remain in their original location may learn about the payoffs in other locations by interacting with newly arrived workers who are likely to have relatively accurate information about their origin. This second-hand information may then influence their future location decisions. To capture local information diffusion, I assume that, after they reach their new destination, agents are able to collect additional information about migration opportunities by communicating with other agents in their location.

In contrast to the rational inattention channel described above, I assume that this second source of information acquisition does not entail any cost, nor any particular decision by agents besides their location decision. Instead, agents naturally form a network with other agents in the same location and communicate their beliefs to all members of the network. I further assume that the local network is complete and that all agents have equal weight in the network.

The standard model of rational learning would require that individuals use Bayes' rule to incorporate any new piece of information into their beliefs. However, in the context of learning in social networks with a large number of other decision-makers, this assumption places unreasonable demands on individuals' cognitive abilities.[20] Here, I sidestep the complex updating by postulating a simple aggregation rule for beliefs in a given location. I set the outcome of the local information sharing to be represented by a log-linear learning rule, resulting in beliefs $\bar{\pi}_{kt}$ about θ_t held by agents in k at the end of t. Since beliefs are the same for all agents i at the end of the period, the prior beliefs for all agents in location j are identical and can be indexed by j. I set

[20]Starting with Degroot (1974), a rich literature has proposed relatively simple functional forms for agents' learning rules, with the objective of capturing the richness of the network interactions while maintaining analytical and computational tractability.

$$\log \bar{\pi}_{kt}(\theta_t) = C_{kt} + \sum_j \sum_s L_{jkt|s} \log \pi_{jt|s}(\theta_t), \tag{2.13}$$

where $L_{jkt|s} = L_{jt-1}\mathbb{E}\left[p_{jkt}(s, \theta_t, \omega_{it})\right]$ is the mass of agents from j in k who received the signal $s_{it} = s$, and C_{kt} is a constant ensuring that $\int_\theta \bar{\pi}_{kt}(\theta)d\theta = 1$.

In addition to its simplicity, this log-linear learning rule has a very intuitive interpretation. First, since all agents have the same weight, a particular posterior belief $\pi_{jt|s}$ will have a larger influence on the final shared belief if more agents in k hold this belief. Second, the log-linear expression implies that the variance of an individual's belief is important. An individual holding beliefs with high variance has little effect on the final beliefs.[21] For instance, in the context of normal beliefs, it is easy to show that the shared beliefs are normal, with a mean equal to a weighted average of each belief's mean, and the weights are inversely proportional to the belief's variance.[22] In Appendix B.1.1, I follow Molavi et al. (2018) and show that the log-linear rule (2.13) can be obtained as the unique aggregation rule under the assumption that agents feature *imperfect recall*. Under this assumption, agents treat the current beliefs of their neighbors as sufficient statistics for all the information available to them, ignoring how or why these opinions were formed. This is a formalization of the idea that real-world individuals do not fully account for the information buried in the entire past history of actions or the complex dynamics of beliefs over social networks.

Sharing information at the local level also implies a simple law of motion for beliefs from one period to the next. Indeed, once information is shared, the beliefs in a given destination no longer depend on where agents come from. The next-period prior beliefs about future productivity can then be expressed as a function of the shared

[21] This contrasts with the heuristic derived by Degroot (1974), under which only the expectation of an individual's belief can determine his influence. For example, an individual with a uniform belief will have no influence on the final belief according to the log-linear learning rule. In contrast, in the model by Degroot (1974), such an individual would be influential as long as his expectation is different from others' expectations.

[22] See Appendix B.1.1.

beliefs at the end of t,

$$\pi_{jt+1}(\theta_{t+1}) = \int_{\theta} \bar{\pi}_{jt}(\theta)\gamma(\theta_{t+1}|\theta)d\theta, \tag{2.14}$$

where $\gamma(\cdot|\theta)$ is the pdf of a normal distribution with mean $\rho\theta$ and variance σ_ξ^2.

2.2.6 Dynamic Rational Inattention Equilibrium

I am ready to define a competitive equilibrium in my economy. A competitive equilibrium consists of a set of information acquisition and mobility strategies such that agents maximize their expected lifetime utility, taking into account the laws of motion for population and beliefs. I denote by J the set of locations, Θ the set of possible productivity vectors, $\mathcal{L} = \{\{L_{jkt}\}_{j,k} | \sum_{j,k} L_{jkt} = \bar{L}, L_{jkt} \geq 0\}$ the set of possible population distributions by previous origins, by S the set of possible signals s, and ΔX the set of distributions over X for any set X. I also define the set $\Omega = \mathcal{L} \times \Delta\Theta \times J \times \mathbb{R}^J$ containing the observed states $\omega_{it} = (L_{t-1}, \pi_{l_{it-1}t}, l_{it-1}, \varepsilon_{it})$.

Definition 1. *Given an initial population distribution L_0 and initial beliefs $\{\pi_{j0}\}_{j\in J}$, an equilibrium is a set of individual information strategies, f, consisting of a system of signal distributions $f_t : \Theta \times \Omega \to \Delta S$, as well as mobility strategies, σ, consisting of a system of mappings $\sigma_t : S \times \Omega \to J$, such that:*

- ***Utility maximization**: Agents solve the problem in (2.10).*
- ***Mobility**: Population evolves as in (2.12).*
- ***Beliefs**: Posterior beliefs are derived from priors according to Bayes' rule (2.7), and are shared locally according (2.13), leading to next-period prior beliefs about future productivity in (2.14).*

I now present a lemma that considerably simplifies the characterization of agents' location choices by allowing me to focus on a special class of information strategies

in which signals correspond directly to actions. The intuition is that it is always optimal to devise an information strategy such that two different signal realizations always lead to different mobility decisions. Receiving distinct signals that would lead to the same decision would be inefficient as information would be acquired but not acted upon. Combining these signals into a single realization would have no effect on the distribution of actions and weakly reduces the information cost. This behavior follows from the convexity of the entropy-based cost function. As a result, the optimal information strategy can be represented as a choice of a distribution of recommendations. Each signal realization essentially reduces to an instruction about which location to choose.[23]

For a given location k, only one signal realization s would lead an agent i to locate in k, so the migration probability $p_{jkt}(s, \theta_t, \omega_{it})$ defined in (2.11) can be expressed as $p_{jkt}(\theta_t, \omega_{it})$ and is equal to the probability that agent i receives the signal s associated to location k. I refer to this probability p as a mobility rule, consisting of a system of distributions over J, for each possible (θ_t, ω_{it}). The following Lemma indicates that instead of solving for the optimal information and mobility strategies in (2.10), I can directly solve for the associated mobility rule p.

Lemma 1. *Any equilibrium strategy (f, σ) that solves the dynamic rational inattention problem in (2.10) generates a choice rule p that solves*

$$\max_{p} \mathbb{E}\left[\sum_{t=1}^{\infty} \delta^t \left(u_{l_{t-1}l_t}(\theta_t, \omega_{it}) - I(\omega_{it})\right)\right], \tag{2.15}$$

[23] In dynamic models, it is necessary to ensure that there is no incentive to select a richer signal structure at time t, for instance to use it for future decisions. Since the information cost is linear in mutual information and agents discount the future, the additive property of entropy ensures that delaying information acquisition never increases the cost, regardless of other information the agent acquires.

where the information cost defined in (2.16) *is expressed as a function of the prior and posterior beliefs* $\pi_{jkt} \equiv \pi_{jt|s_t=k}$*:*

$$I(\omega_{it}) = \lambda_{l_{it-1}} \left(H(\pi_{jt}) - \sum_k q_{jkt}(\omega_{it}) H(\pi_{jkt}) \right) \qquad \forall j \tag{2.16}$$

and $q_{jkt}(\omega_{it}) = \int_\theta p_{jkt}(\theta, \omega_{it}) \pi_{jt}(\theta) d\theta$ *is the ex-ante probability of receiving the recommendation to move to* k*. Population and beliefs follow the laws of motion* (2.12) – (2.14)*.*

Proof. See Appendix B.1.1. □

Accordingly, I will dispense with the signals s_t, replacing them with actions l_t, and will refer to any rule p solving (2.15), and the laws of motion of population and beliefs, as a solution to the dynamic rational inattention equilibrium.

Proposition 1. *There exists a solution to the dynamic rational inattention equilibrium.*

Proof. See Appendix B.1.1. □

Proposition 1 extends the existence result derived in Steiner et al. (2017) to the case of a non-finite state space with a continuum of agents and endogenous payoffs by ensuring that the strategy space is compact, and that the boundedness of payoffs together with discounting ensure that agents' objective functions are continuous in their strategies.

2.2.7 Stochastic Steady-State Equilibrium

I now introduce a number of assumptions that will help deliver a tractable solution to the dynamic rational inattention equilibrium in stochastic steady state. A stochastic steady state consists of a mobility rule, $p(\cdot)$, and beliefs, $\pi(\cdot)$, that are time-invariant,

in the sense that they always map a given set of states to the same actions and probability distributions, respectively. In the stochastic steady state, there is still variation in mobility flows and earnings each period as new productivity vectors are realized. However, population and beliefs in each location evolve within an ergodic distribution, and we can drop the t subscript for the mobility rule and beliefs.

Definition 2. *A stochastic steady state equilibrium is a mobility rule* $p : \Theta \times \Omega \to J$*, as well as beliefs* $\pi : \Omega \to \Delta\Theta$*, such that* p *is a solution to* (2.15)*, while population and beliefs follow the laws of motion* (2.12) – (2.14) *and satisfy* $p_{jkt}(\theta, \omega) = p_{jkt+1}(\theta, \omega)$*,* $\pi_{jt}(\theta|\omega) = \pi_{jt+1}(\theta|\omega)$*.*

Note that the state variables that migration decisions depend on contain the prior beliefs at the beginning of the period. Indeed, even if we consider one location and two different time periods at which the productivity, population distribution and the agents' preferences are identical but prior beliefs are different, we may still expect agents to acquire different amounts of information, resulting in different beliefs and mobility decisions. However, as I discuss in Appendix B.1.1, under the condition that the variance of productivity process is not too large, in stochastic steady state the prior beliefs $\pi_j(\theta_t|\omega_{it})$ can be expressed as a function of only the population distribution L_{t-1}, $\pi_j(\theta_t|L_{t-1})$. As a result, the effective states that determine agents' mobility decisions are then $(\theta_t, L_{t-1}, \varepsilon_{it})$. This property of beliefs is specific to rational inattention problems and results from the property of *locally invariant posteriors* shown in the context of a static model by Caplin and Dean (2015).[24]

[24]The *locally invariant posteriors* property states that agents with different priors will locally choose the same posterior beliefs. If this property holds in our dynamic context, agents who move from j to k always have the same posterior beliefs, so that the shared beliefs in k in some period t only depend on the composition of agents at t, and not on their prior beliefs. Imposing that the variance of the productivity process is sufficiently small ensures that the population composition of a location, and the resulting priors, remain in the set leading to the same posterior beliefs. To check that this property holds in practice, I solve the model with 10 locations, allowing prior beliefs to vary with (L_{t-1}, L_{t-2}) and posteriors to vary with L_{t-1}, and show that for reasonable values of fundamentals and productivity process parameters, the priors do not vary with L_{t-2}.

I am now ready to characterize the solution to the stochastic steady-state. As was shown by Matejka and McKay (2015) in a static framework, and extended by Steiner et al. (2017) to a dynamic setting, the specific form of the information cost (2.9) based on the Shannon entropy leads to a modified logit form for the decision rule. The following Proposition shows that the mobility rule features a similar logit structure in the stochastic steady state.

Proposition 2. *In the stochastic steady state, for each agent located in j at time $t-1$, the optimal mobility rule $p_{jk}(\theta_t, L_{t-1}, \varepsilon_{it})$ can be expressed as*

$$p_{jk}(\theta_t, L_{t-1}, \varepsilon_{it}) = \frac{q_{jk}(L_{t-1}, \varepsilon_{it}) \exp\left(u_{jk}(\theta_t, L_{t-1}, \varepsilon_{it}) + \delta \bar{V}_k(\theta_t, L_{t-1})\right)^{1/\lambda_j}}{\sum_l q_{jl}(L_{t-1}, \varepsilon_{it}) \exp\left(u_{jl}(\theta_t, L_{t-1}, \varepsilon_{it}) + \delta \bar{V}_l(\theta_t, L_{t-1})\right)^{1/\lambda_j}} \tag{2.17}$$

where $q_{jk}(L_{t-1}, \varepsilon_{it}) = \int_\theta p_{jk}(\theta, L_{t-1}, \varepsilon_{it}) \pi_j(\theta | L_{t-1}) d\theta$ and we define the expected future value as $\bar{V}_k(\theta_t, L_{t-1}) = \mathbb{E}\left[V_k(\theta_{t+1}, L_t, \varepsilon_{it+1}) | \theta_t, L_{t-1}\right]$. The continuation payoffs solve

$$V_j(\theta_t, L_{t-1}, \varepsilon_{it}) = \lambda_j \log\left(\sum_l q_{jl}(L_{t-1}, \varepsilon_{it}) \exp\left(u_{jl}(\theta_t, L_{t-1}, \varepsilon_{it}) + \delta \bar{V}_l(\theta_t, L_{t-1})\right)^{1/\lambda_j}\right), \tag{2.18}$$

and population and beliefs follow the laws of motion (2.12) – (2.14).

Proof. See Appendix B.1.1. □

Conditional on a productivity vector θ_t, a population distribution L_{t-1} and preference shocks ε_{it}, agent i's decision is stochastic because it depends on which signal realization she obtains. If information was complete, an agent in j would observe the payoffs $u_{jk} + \delta \bar{V}_k$ for all k under the states $(\theta_t, L_{t-1}, \varepsilon_{it})$, choose the destination k^* offering the highest payoffs, so that the mobility rule p_{jk} would be an indicator function equal to 1 for this particular k^*, zero for all other destinations. This is indeed the limit of the mobility rule (2.17) as $\lambda_j \to 0$. When information is costly and $\lambda_j > 0$, the expression (2.17) indicates that a destination k that offers higher payoffs

$u_{jk} + \delta \bar{V}_k$ under the states $(\theta_t, L_{t-1}, \varepsilon_{it})$ has a higher probability of being selected after the agent has acquired information. However, the expression also contains an endogenous bilateral shifter in the form of the ex-ante moving probability q_{jk}. The shifter predisposes agents in j to move to k, by a magnitude that depends on the prior beliefs held by agents in j about payoffs in k. If the expected probability of moving to k is large according to the current population composition, then moving to k is more likely, irrespective of the current realization of productivity θ_t. This endogenous predisposition is a sufficient statistic expressing the magnitude with which information frictions favor some migration decisions relative to others. If a location k offers high payoffs for agents in j, either because of high productivity, amenities or low migration costs, this translates into higher average migration flows between j and k and therefore into a higher predisposition for this transition.

The expected value of being located in some location k in (2.18) solves a Bellman equation akin to dynamic logit models. The value of being in some location j is increasing with the payoffs that can be expected from moving to other locations. If information was complete, the value of being in location j when the state is $(\theta_t, L_{t-1}, \varepsilon_{it})$ would simply be the payoff $u_{jk^*} + \delta \bar{V}_{k^*}$ in the optimal location k^*. When information is incomplete, the agent may move to any location l with some probability depending on the information they obtain. The payoffs of each potential destination l are weighted by the ex-ante probability that agents in j move to l. The contribution of information costs defined in (2.16) is summarized by the endogenous predispositions q_{jk} and the exponent λ_j.

At this stage, mobility decisions have a different expression for every preference draw. In order to express the total bilateral mobility flows for each productivity and population (θ_t, L_{t-1}), I need to aggregate these individual decisions by integrating over preference shocks ε_{it}. Moreover, even if the mobility rule is only a function of (θ_t, L_{t-1}), the number of state variables is large even for a limited number of loca-

tions. As a result, I approximate the equilibrium belief distribution when I solve for the mobility rule. I choose to approximate the belief distribution by the conjugate of a Type 1 extreme value (EV1) distribution, which leads to a closed-form aggregation over preference shocks. In addition, I assume that preference shocks are also drawn from this class of distributions.[25] The main property of the conjugate EV1 distribution is that if a random variable X is drawn from a EV1 distribution and another random variable Y is drawn from a conjugate EV1 distribution, then $Y + X$ is a random variable distributed as EV1.[26]

Assumption 1. *Equilibrium marginal beliefs $\pi_j(\theta_{kt}|L_{t-1})$ can be approximated by independent conjugate EV1 distributions with mean $\mu_{jk}(L_{t-1})$ and same variance σ_j^2.*

Assumption 2. *Preference shocks ε_{ikt} are drawn from independent conjugate EV1 distributions with mean zero and dispersion ν.*

Although Assumption 2 is similar to the usual assumption that preference shocks are drawn from Type I extreme value distributions, Assumption 1 imposes restrictions on the equilibrium behavior of agents by constraining beliefs to belong to a particular distribution and to be independent across locations. Before presenting the expression for the migration probability $\bar{p}_{jk}(\theta_t, L_{t-1}) = \mathbb{E}\left[p_{jk}(\theta_t, L_{t-1}, \varepsilon_{it})\right]$ integrated over preference shocks ε_{it}, I denote by $\bar{u}_{jk}$ the average utility flow $\mathbb{E}\left[u_{ijk}\right]$ integrated over ε_{it} and define $\bar{v}_{jk}(\theta_t, L_{t-1})$ as the expected value of moving from j to k if the productivity and population are (θ_t, L_{t-1}):

$$\bar{v}_{jk}(\theta_t, L_{t-1}) = \bar{u}_{jk}(\theta_t, L_{t-1}) + \delta \bar{V}_k(\theta_t, L_{t-1}).$$

[25] Brown and Jeon (2019) also imposed that beliefs and preferences follow a conjugate EV1 distribution in their static model of optimal health insurance choice. Dasgupta and Mondria (2014) imposed that productivity shocks follow a particular type of conjugate EV1, the Cardell distribution, in their model of international trade.

[26] The relation between the conjugate EV1 distribution and the EV1 distribution is displayed in Appendix B.5. Qualitatively, the two distributions are very similar.

I also define the expectational error, $\chi_{jk}(\theta_t, L_{t-1}) = \mu_{jk}(L_{t-1}) - \theta_{kt}$, as the difference between the expected value of θ_{kt} according to agents in location j at time t, and its actual realization θ_{kt}.

Proposition 3. *Under Assumptions 1 and 2, the average mobility rule in the presence of information frictions and preference heterogeneity is given by*

$$\bar{p}_{jk}(\theta_t, L_{t-1}) = \frac{\exp\left(\eta_j \chi_{jk}(\theta_t, L_{t-1}) + \bar{v}_{jk}(\theta_t, L_{t-1})\right)^{\frac{1}{\phi_j}}}{\sum_l \exp\left(\eta_j \chi_{jl}(\theta_t, L_{t-1}) + \bar{v}_{jl}(\theta_t, L_{t-1})\right)^{\frac{1}{\phi_j}}}, \tag{2.19}$$

where $\chi_{jk}(\theta_t, L_{t-1}) = \mu_{jk}(L_{t-1}) - \theta_{kt}$ is the expectational error made by agents in j about θ_{kt}, while the continuation payoffs solve

$$V_j(\theta_t, L_{t-1}) = \phi_j \log\left(\sum_l \exp\left(\eta_j \chi_{jl}(\theta_t, L_{t-1}) + \bar{v}_{jl}(\theta_t, L_{t-1})\right)^{\frac{1}{\phi_j}}\right), \tag{2.20}$$

with $\phi_j = \nu\left(1 + \lambda_j^2 (1-\eta_j)^2\right)^{1/2}$ and $\eta_j = \left(1 + \frac{6\sigma_j^2}{\pi^2 \lambda_j^2 \nu^2}\right)^{-1/2} \in (0,1)$, and π is the constant $\pi = 3.1415...$

When the information cost tends to zero, $\mu_{jk}(L_{t-1}) \to \theta_{kt}$ and $\phi_j \to \nu$ so that the model reduces to a preference-based migration model:

$$\bar{p}_{jk}(\theta_t, L_{t-1}) \to_{\lambda_j \to 0} \frac{\exp\left(\bar{v}_{jk}(\theta_t, L_{t-1})\right)^{1/\nu}}{\sum_l \exp\left(\bar{v}_{jl}(\theta_t, L_{t-1})\right)^{1/\nu}}. \tag{2.21}$$

When the dispersion of preferences ν tends to zero, the solution becomes

$$\bar{p}_{jk}(\theta_t, L_{t-1}) = \frac{\exp\left(\rho_j \chi_{jk}(\theta_t, L_{t-1}) + \bar{v}_{jk}(\theta_t, L_{t-1})\right)^{\frac{1}{\psi_j}}}{\sum_l \exp\left(\rho_j \chi_{jl}(\theta_t, L_{t-1}) + \bar{v}_{jl}(\theta_t, L_{t-1})\right)^{\frac{1}{\psi_j}}}, \tag{2.22}$$

where $\psi_j = \lambda_j (1 - \rho_j)$ and $\rho_j = \left(1 + \frac{6\sigma_j^2}{\pi^2 \lambda_j^2}\right)^{-1/2} \in (0,1)$.

Proof. See Appendix B.1.1. □

The role of information frictions λ_j in altering the responsiveness of migration to local shocks is apparent in (2.19). The elasticity of migration with respect to observed components of $\bar{u}_{jk}$ is $1/\phi_j$, whereas the elasticity of migration with respect to θ_t is smaller and equal to $(1-\eta_j)/\phi_j$, after including the contribution of θ_t to the expectational error χ_{jk}. Since $\eta_j \in (0,1)$, migration flows are less responsive to variations in θ_t than they are to variations in observed payoffs, precisely because agents must incur a cost to learn about the value of θ_t and therefore choose to only imperfectly observe θ_t. When information costs λ_j become very large, $\eta_j \to 1$ and the elasticity of migration with respect to θ_t is zero. Conversely, when information costs are zero, prior beliefs about θ_t become perfectly accurate and $\mu_{jk}(L_{t-1})$ equals θ_{kt}. As shown by (2.21), this implies that the elasticity of migration with respect to all components of the payoffs $\bar{v}_{jk}$, including θ_t, equals $1/\nu$. Therefore, the role of information frictions is identified by the *differential* responsiveness of migration flows with respect to unobserved (θ_t) and observed (L_{t-1}) determinants of payoffs. In Section 2.4.2, I propose a strategy for estimating the elasticities $(1/\phi_{jt}, (1-\eta_{jt})/\phi_{jt})$ that exploit the mapping from these elasticities to the information costs and preference heterogeneity (λ_j, ν).

In addition to affecting the responsiveness of migration flows to fluctuations in payoffs, information frictions affect the levels of migration flows. This mechanism shows up in (2.19) as the role of mean prior beliefs $\mu_{jk}(L_{t-1})$ in shifting the level of bilateral migration flows. From the expression of shared beliefs (2.13), we can see for example that the beliefs held by agents in location k about the productivity in k are likely to be biased upwards. Agents who decided to move to k all received signals that were favorable to k, leading to a consensus that attaches a high probability to large values of θ_{kt}. This high mean prior belief about θ_{kt} translates into a higher probability of staying in location k in the next period. More generally, destinations that agents are optimistic about will be favored over other locations, creating the same

endogenous predispositions towards some locations as was discussed in reference to (2.17). As I will discuss in Section 2.4.4, failing to account for information frictions will overestimate the actual bilateral migration costs κ_{jk}. The dependence of mobility decisions on mean beliefs also has implications for mobility patterns. As I will argue in Section 2.5, the covariance of mean beliefs with the realized productivity varies across pairs of locations and delivers migration patterns that are in line with observed flows. In particular, agents' beliefs are less responsive to productivity shocks in distant locations, both because they are less likely to interact with well-informed people from these locations, and because they individually gather less information about them. This leads to a decreasing migration elasticity with distance.

2.3 Data Sources

In this section, I present my main sources of data and provide descriptive statistics on migration patterns in Brazil. I provide more information about how I constructed my sample in Appendix B.2.

The main source of data is the Relação Anual de Informações Sociais (RAIS), which is collected annually by the Brazilian Ministry of Labor and contains matched employer-employee information for every formally employed worker in Brazil.[27] It includes demographic, occupational and income characteristics for employees, with individual identifiers so that workers can be followed from year to year. I use 15 consecutive years of data, corresponding to the period between 2000 and 2014. RAIS also includes the geographic location of each employment contract at the municipality level. For every formal job and year, I exploit information on the duration of the job

[27] In Brazil, informal and self-employed workers constitute a large fraction of the labor force, peaking at 45% in 2000 (Bilbiie et al., 2012). Consequently, my analysis leaves out a significant fraction of the labor force. Workers may consider transitions into and out of the formal sector as an alternative to geographic mobility. To limit the prevalence of these transitions, I restrict my analysis to workers with a high attachment to the formal sector by focusing on workers spending at least seven consecutive years in the formal sector over the sample period.

spell, the average monthly wage, the number of hours stipulated in the contract, and certain characteristics of the plant where the worker is employed.[28] Specifically, I use information on the micro and mesoregion in which the plant is located.[29]

Workers in the sample often appear to be performing multiple different jobs in the same year. If a worker has more than one job in a year, I assign them to the job they held for longest during that year. However, to determine the total labor income of a worker in a year, I add labor income from all jobs that worker held.

The data contains no information on the residence of a worker before their first job in the formal sector. Consequently, in the analysis, I focus on the migration decisions of workers and do not model the decision to enter the labor force or acquire college education. For this reason, I limit the data to workers that are over 25 years of age, since, for the majority of the population, education decisions are taken before this age. Similarly, I do not model the retirement decision of workers and, consequently, I limit the data workers below 65 years of age.

Besides the information on workers' labor market histories contained in the RAIS database, I also use information on the population of each municipality in each year between 2000 and 2014 from the population census collected by *Instituto Brasileiro de Geografia e Estatistica* (IBGE). I compute population by mesoregion by aggregating the population of all municipalities included in the corresponding mesoregion.

Finally, I collect information on the degree of internet penetration by microregion. Specifically, from the *Agência Nacional de Telecomunicações*, ANATEL, a government agency in charge of regulating and supervising telecommunications in Brazil, I obtain information on the number of broadband connections by municipality and year

[28] In the analysis of migration patterns at the individual level described in section 2.5.1, I also exploit the 2-digit occupation (according to the *Classificação Brasileira de Ocupações*, CBO), the 2-digit industry of production of the establishment for the main job spell (according to the *Classificação Nacional de Atividades Econômicas*, CNAE), as well as information on the workers' gender, age, and level of education.

[29] Brazilian microregions are groups of municipalities that span the entirety of the Brazilian territory and are the closest equivalent to commuting zones. During our sample period, there were 558 microregions which are grouped into 137 mesoregions.

between 2007 and 2014. I use the population data obtained from the population census to construct a municipality-specific measure of the number of broadband connections per capita, and again use the municipality-level population data to construct an equivalent variable at the mesoregion level. As I do not have access to information on the number of broadband internet connections for the years 2000 to 2006 and the available data indicates that the overall number of broadband connections per inhabitant is less than 1% in 2007, I assume that the number of broadband connections per capita equals zero in every mesoregion before 2007.

The resulting dataset includes 372,454,979 worker-year pairs that correspond to 45,958,805 workers. According to the RAIS data, the 137 mesoregions had on average 250,000 legal workers in 2014 (the median microregion had close to 62,000 legal workers). The mesoregions with the highest labor force are located in the South and along the coast. The average yearly rate of migration across mesoregions is 3.4%.

2.4 Estimating the Model

In this section, I structurally estimate my model, guided by the mobility rule derived in the previous section. I estimate the model separately over two time periods: 2000-2007 and 2008-2014. I will treat each of these periods as steady states, and compare the parameters, in particular information costs, during each period. I follow a two-step procedure for each period: first, I estimate the production parameters that can be inferred directly from the data without simulating the model. Second, I use the gravity equation of migration predicted by the model to estimate the remaining parameters consisting of the amenities, migration costs, preference heterogeneity, and information costs. The structural estimation uses data from RAIS described in Section 2.3. I construct yearly bilateral migration flows between each of the 137 mesoregions in

Brazil, and average wages in each region and year, from individual records of earnings and location each year.

2.4.1 Step 1: Production parameters

I begin by estimating $J + 3$ parameters that can be obtained directly from the data. This step only imposes assumptions on the production side of my model. These are the parameters governing the decreasing returns in labor, $1 - \alpha$, the persistence and volatility of the productivity process, (ρ, σ_ξ^2), and the regional baseline productivities A_k.

I first set the Cobb-Douglas share of labor, $1 - \alpha$, used in production equal to the average labor share computed from national statistics over each period.[30] The labor share, $1 - \alpha$, in Brazil is stable between 2000 and 2014 at a value of 0.60.

I then estimate the parameters of the productivity process. The Brazilian economy is growing over the sample period. The total employment recorded in RAIS is also growing each year, both from demographic change and increased transitions to the formal labor market. To be able to interpret the data as closely as possible to a steady state with constant total population, I project wages on year fixed effects and normalize all population stocks so that the total population in the economy is constant at its 2000 level over the sample. I then estimate the AR(1) process associated with the observed wages after substituting year fixed effects and adjusting for population, $\tilde{w}_{kt} = \log w_{kt} + \alpha \log L_{kt} - \log(1-\alpha)$. From the first order condition determining wages (2.4) and the evolution of productivity (2.3), the adjusted wages can be expressed as

$$\tilde{w}_{kt} \equiv \log w_{kt} + \alpha \log L_{kt} - \log(1 - \alpha) = A_k + \theta_{kt}. \tag{2.23}$$

[30] See Restrepo-Echavarria and Reinbold (2018). The data available in RAIS only contains payments made to formally employed workers and does not report information on value added. The labor share is computed from the Penn World Tables.

The baseline productivities are therefore recovered as the average of $\tilde{w}_{kt}$ over the period, since $\mathbb{E}[\theta_{kt}] = 0$. The persistence of the productivity process ρ is equal to the covariance of $\tilde{w}_{kt}$ and $\tilde{w}_{kt-1}$ divided by the variance of $\tilde{w}_{kt}$. The volatility of the productivity process is then computed as $(1-\rho^2)\text{Var}(\tilde{w}_{kt})$. Finally, I recover the productivity shocks θ_{kt} implied from the observed $\tilde{w}_{kt}$ and estimated productivities A_k as the residuals from (2.23).[31] As reported in Table 2.1, the estimated productivities are on average equal to 4.49 with a standard deviation of 0.87 across regions between 2000 and 2007. For 2008–2014, the mean and standard deviations are 4.72 and 0.83 respectively. The persistence of the productivity process ρ is 0.76 in the first period and 0.69 in the second period. The volatility of the productivity process σ_ξ^2 is 0.33 in the first period and 0.38 in the second period.

2.4.2 Step 2: Simulated Method of Moments

In the second stage, I use the method of simulated moments to estimate the remaining parameters $\vartheta = (\nu, \lambda_j, B_j, \kappa_{jk})$, which consist of the dispersion of idiosyncratic preferences, information costs, amenities, and migration costs. The total number of parameters to be estimated in this step is large: there are as many as $2J+(J-1)^2 = 18770$ of them.[32] A grid search over the parameter space is therefore not practical. To circumvent this issue, I develop an iterative algorithm that updates the parameter guesses in a simple intuitive way and delivers fast convergence. First, I use the mobility rule in (2.19) to obtain a regression equation predicted by the model. This regression equation offers moments that I target to identify the parameters of interest. I then simulate the model given parameter estimates from the first step and guesses for the

[31] Since the production specification does not account for the use of capital or other inputs, any variation in observed wages that is caused by a shock to these variables is interpreted as a productivity shock.

[32] There are J information costs, $J-1$ amenities since one can be normalized to 0, $(J-1)^2$ migration costs since I normalize $\kappa_{jj} = 0$ for all j, and the preference heterogeneity ν.

parameters to be estimated. I run the predicted regression in the model and update the parameter guesses using the estimated coefficients of the regression.

To obtain the regression equation at the core of my estimation, I exploit the gravity structure of the model. Denote by $\bar{p}_{jkt}$ the migration share between j and k conditional on the productivity and population at time t. Taking the log of the migration share $\bar{p}_{jkt}$ divided by the share of stayers in j, $\bar{p}_{jjt}$, and using the expression of $\bar{u}_{jk}$ in (2.5) yields the following expression:

$$\log \frac{\bar{p}_{jkt}}{\bar{p}_{jjt}} = \frac{\eta_j}{\phi_j} \left(\chi_{jkt} - \chi_{jjt}\right) + \frac{1}{\phi_j} \left(\log \frac{w_{kt}}{w_{jt}} + D_{jk} + \delta \left(V_{kt+1} - V_{jt+1}\right) \right) + e_{jkt}, \quad (2.24)$$

where $D_{jk} = B_k - B_j - \kappa_{jk}$ is a composite bilateral "resistance" term combining the contributions of amenity differences between regions and migration costs. The error term $e_{jkt} = \left(\zeta_{jjt} - \zeta_{jkt}\right)/\phi_j$ is composed of expectational errors about future values $\zeta_{jkt} = \delta \left(V_{kt+1} - E_{j,t} V_{kt+1}\right)$. Equation (2.24) has a "gravity" structure in that the magnitude of bilateral flows is increasing in the wage (and future value) differentials between any two regions, and is decreasing in the "distance" between regions captured by D_{jk}. The new contributor to this gravity structure is the "optimism differential" that agents express towards the destination relative to their origin, and is represented by the difference between the expectational errors made by agents in j about θ_{kt} and θ_{jt}, $(\chi_{jkt} - \chi_{jjt})$.

Since the error term e_{jkt} only arises because of the irreducible uncertainty about future productivity θ_{t+1} due to the innovation of the AR(1) process, it is orthogonal to all other regressors. As a result, the coefficients $\hat{\beta}$ obtained from estimating (2.24) by ordinary least squares (OLS) are the method of moments estimator for the moment condition associated to the orthogonality of the regressors and the error term. There is a unique mapping between the coefficients of the regression $\beta = (\eta_j/\phi_j, 1/\phi_j, \{D_{jk}/\phi_j\})$, and the parameters of interest $\vartheta = (\nu, \lambda_j, B_j, \kappa_{jk})$. In-

tuitively, η_j and ϕ_j together identify ν and λ_j, while projecting D_{jk} on origin and destination fixed effects allows me to recover amenities, B_j, and fixed costs, κ_{jk}, after normalizing $\kappa_{jj} = 0$ for all j.[33]

The estimation algorithm I employ relies on a standard fixed point iteration on the OLS equation (2.24). A satisfactory set of parameters ϑ should deliver beliefs and values such that the estimates $\hat{\beta}$ from (2.24) map exactly to $\hat{\vartheta}(\hat{\beta}) = \vartheta$. First, I guess initial values $\vartheta^0 = (\nu^{(0)}, \lambda_j^{(0)}, B_j^{(0)}, \kappa_{jk}^{(0)})$ for the parameters to be estimated.[34] Second, I simulate the model using the production parameters $(\alpha, \rho, \sigma_\xi^2)$ estimated in the first block, and the current parameters $\vartheta^{(0)}$. This delivers mean belief functions $\mu^{(0)}(\cdot)$ and values $V^{(0)}(\cdot)$. Third, I evaluate beliefs and values at observed populations L_{t-1}^{obs} and recovered θ_t^{obs}, and estimate (2.24) by OLS.[35] Fourth, from the estimated η_j and ϕ_j and fixed effects, I update the parameters to $\vartheta^{(1)}$ and return to the second step using $\vartheta^{(1)}$. I keep iterating until my estimator converges to a fixed point ϑ^*. By construction, this procedure minimizes the method of moment objective associated to the moment conditions of the OLS regression (2.24). I describe the iterative algorithm in more detail in Appendix B.3.1. Despite the absence of theoretical results on the convergence of this algorithm, I find that in practice, it converges quickly and always to the same solution ϑ^* for different initial values $\vartheta^{(0)}$.

I estimate the model over the two different periods, 2000–2007 and 2008–2014, and report the results in Table 2.1. I describe the information costs in Section 2.4.3. The estimated dispersion of preferences is equal to 2.31 in the first period and 2.62 in the second period. The dispersion in preferences appears to have increased over time as information costs decreased. This reflects the fact that migration flows have

[33] The mapping between (η_j, ϕ_j) and (ν, λ_j) is not direct as it depends on the variance of beliefs σ_j, itself a function of all parameters. See Appendix B.3.1 for more details.

[34] To obtain reasonable initial guesses, I first run (2.24) omitting the mean beliefs and future values.

[35] In the data, out of the $(J-1)^2 = 136^2 = 18,496$ migration trips that could be undertaken between any two mesoregions, only $10,382$ of them have positive flows recorded more than once over the 15 years of data. This implies that no fixed bilateral resistance term can be estimated for these pairs, and I assign them a prohibitive migration cost.

Table 2.1: Estimated Parameters

Parameter	Related Moment	Statistic	Time Period	
			2000-2007	2008-2014
α	Inverse Labor Share	Value	0.41	0.40
ρ	Persistence of Wages	Value	0.76	0.69
σ_ξ^2	Volatility of Wages	Value	0.33	0.38
ν	Migration Elast. wrt wages	Value	2.31	2.62
λ_j	Migration Elast. wrt Productivity	Mean	3.11	2.23
A_j	Average Wages	Std Dev	0.87	0.83
B_j	Average Population	Std Dev	0.83	0.91
κ_{jk}	Average Migration Flows	Mean	0.56	0.47

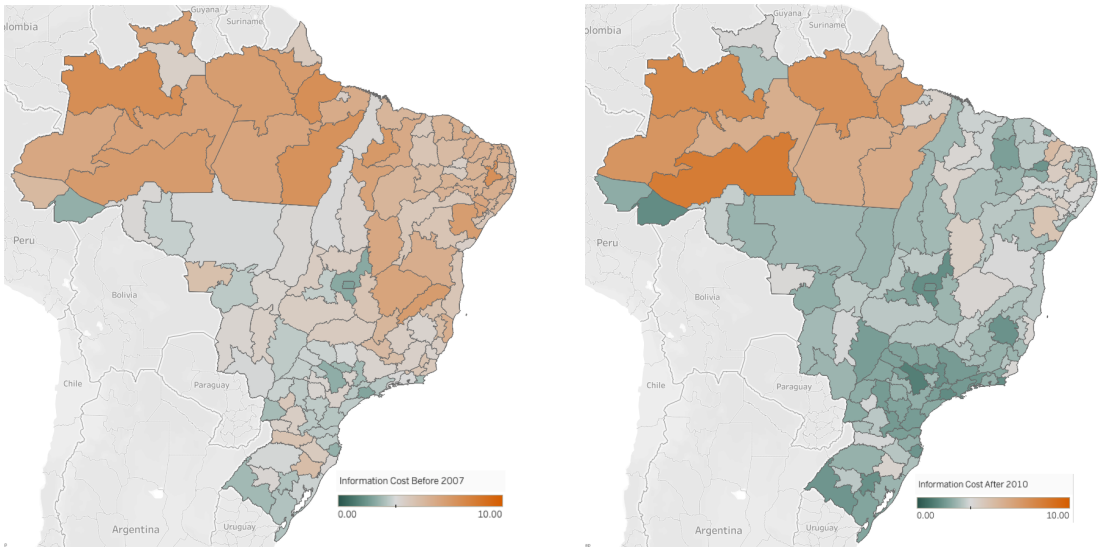
Estimated values of the parameters of the production process, regional baseline productivities, amenities, and information costs, as well as bilateral migration costs.

become less responsive to wages overall, but have become relatively more responsive to unobserved productivity shocks. Finally, I separate the amenities from the migration costs by projecting the estimated D_{jk} on origin and destination fixed effects. I normalize the lowest estimates of amenity across regions to zero, and obtain that the standard deviation is 0.83 in the first period and 0.91 in the second period. The average migration costs are 0.56 in the first period and 0.47 in the second period. The decline in migration costs between the two periods is consistent with the increase in migration flows between the two periods.[36] Overall, parameter values are fairly similar over the two periods.

2.4.3 Information Costs and Internet Access

Information costs are on average equal to 3.11 in the first period and 2.23 in the second period. The standard deviation in the information costs is 1.43 in the first period and 1.32 in the second period. Figure 2.2a and Figure 2.2b display the estimated information costs λ_j for each of the 137 mesoregions for the first and second periods. The information costs appear strongly correlated with the economic devel-

[36] See Appendix B.2.2 for a description of the yearly migration rates.



(a) Information Costs 2000-2007 (b) Information Costs 2008-2014

opment of regions, with the lowest costs of information in the most densely populated and richest regions of the South, including the metropolitan areas of São Paulo, Rio de Janeiro, and Brasilia. In contrast, the less developed North-Eastern regions and remote regions in the Amazon appear to have a much higher cost of information. It is worth emphasizing that these information costs were recovered as fixed-slope coefficients with no parametric assumption regarding their correlation with any observable variable. The lower information costs obtained for Southern regions result from the fact that workers in these regions appear to be relatively more responsive to the unobserved productivity shocks than workers in other regions.

Since these information costs are novel parameters that have not yet been estimated in the literature, I now investigate their variation along observable regional characteristics. From inspecting Figure 2.2a and Figure 2.2b, these information costs appear to be correlated with population density and average income. One intuitive shifter for the technological cost of information acquisition is the availability of internet in the region. There may also be persistent determinants of the information costs at the region level, such as geographic accessibility. Motivated by these remarks, I

Table 2.2: Determinants of Information Costs

	OLS	IV
Internet Connections / Inhabitants	-0.971^a	-0.831^b
	(0.321)	(0.461)
Log Income	-0.824^a	-0.793^a
	(0.241)	(0.262)
Log Pop. Density	-0.310^c	-0.323^c
	(0.190)	(0.202)
Observations	137	137
R^2	0.219	0.192

a denotes 1% significance, b denotes 5% significance, c denotes 10% significance. In parenthesis, I report standard errors.

propose a simple parameterization of information costs:

$$\lambda_{jt} = \ell_1 int_{jt} + \ell_2 \log w_{jt} + \ell_3 \log popdens_{jt} + \varsigma_j + u_{jt}, \tag{2.25}$$

where $t \in \{1, 2\}$ corresponds to each of the two periods 2000–2007 and 2008–2014, int_{jt} is the average fraction of residents of j with an active internet connection over the years corresponding to the period $t = 2$, and zero for the period $t = 1$, $\log w_{jt}$ is the log average wage in region t over the years corresponding to the period t, $\log popdens_{jt}$ is the average log of the population density in region j during each period, ς_j captures unobserved determinants of the information costs that are constant across the two periods, and u_{jt} captures the unobserved time varying determinants of the information costs.

I estimate (2.25) in first differences by OLS and report the results in the first column of Table 2.2. The coefficient on internet access is large and significant at the 1% level, and indicates that conditional on income and population density, increasing the fraction of residents with an internet connections from zero to one is associated with a decline in the information cost by 0.97 units. The coefficient on log income,

also significant at the 1% level, implies that a 1% increase in local income is associated with a decrease in the information cost by 0.00824 units. The coefficient on population density is only significant at the 10% level and reveals that a 1% increase in population density is associated with a 0.0031 decrease in the information cost.

In the analysis of counterfactual exercises described in Section 2.6, I am interested in reproducing the plausible decrease in information costs brought about by the expansion of internet access at the local level. From inspecting (2.25), one may suspect that there could be unobserved time varying factors in the error u_{jt} that are correlated with the fraction of households with an internet connection. For example, changes in local public spending on transportation or communication infrastructure may have directly reduced the cost of information, and facilitated internet expansion. These omitted variables may lead to an upward bias in the estimation of the causal effect of internet penetration on the information cost.

To circumvent this issue, I instrument the fraction of residents with an internet connection with a dummy variable equal to 1 if the region is located less than 250 km away from a backbone cable.[37] Only agents residing in the vicinity of a backbone cable can expect to have access to high speed internet. Importantly, the geographic coverage of these backbones, including the ones deployed over the period 2008–2014, follows other infrastructure that existed prior to 2008. This provides a plausible source of variation for the extent of internet penetration that is unlikely to be affected by later changes in local economic conditions. To construct the instrument, I follow Tian (2019); see Appendix B.3.2. I report the results from the instrumental variable regression in column 2 of Table 2.2. The magnitude of the coefficient of internet connections per resident is slightly lower, but remains large and significant at the 5% level.

[37] "Backbones" are national trunk infrastructure that brings traffic from international submarine cables in coastal regions to inland parts of the country. Backbones consist of high-capacity fiber optic cables. The quality of broadband connection decreases exponentially with distance to the backbone, and I follow Tian (2019) in setting the threshold of 250 km.

2.4.4 Migration Costs With and Without Information Frictions

I now illustrate the implications of my model for the magnitude of migration costs and preference heterogeneity. Since information frictions are a source of both endogenous migration costs and limited migration elasticity, smaller exogenous migration costs κ_{jk} and preference heterogeneity ν will be necessary to explain the observed migration patterns. I confirm this prediction by estimating the model under the assumption that there are no information frictions, so $\lambda_j = 0$ for every region j. This corresponds to the mobility rule (2.21).

Since there are no unobserved expectational errors to control for, it is possible to devise a direct estimation strategy without simulating the model. This method relies on renewal actions and compares migration paths that visit the same locations at t and $t+2$ but differ at $t+1$ (Artuç et al., 2010; Traiberman, 2019; Caliendo et al., 2018). From $t+2$ onwards, these paths offer the same continuation value, so the difference in the payoffs they offer can be expressed as a function of the wages in the visited regions at time $t+1$. If I select the origin destination at t to be j, the location at $t+1$ to be either j or k, and the destination at $t+2$ to be k, I can write the estimation equation

$$\Lambda_{jkt} = \frac{1}{\nu}\left(\Delta \log w_{jkt} + D_{jk} + e_{jkt}\right), \tag{2.26}$$

where $D_{jk} = B_k - B_j - \kappa_{jk}$ and $\Lambda_{jkt} = \log\left(\frac{p_{jkt}p^{\delta}_{kkt+1}}{p_{jjt}p^{\delta}_{jkt+1}}\right)$ is the relative discounted probability of the paths $j \to k \to k$ and $j \to j \to k$. The residuals χ_{jkt} are a collection of expectational errors that are orthogonal to the regressors:

$$\chi_{jkt} = \zeta_{jjt} - \zeta_{jkt} + \delta(\zeta_{jkt+1} - \zeta_{kkt+1}), \qquad \zeta_{jkt} = \delta\left(V_{kt+1} - E_{j,t}V_{kt+1}\right).$$

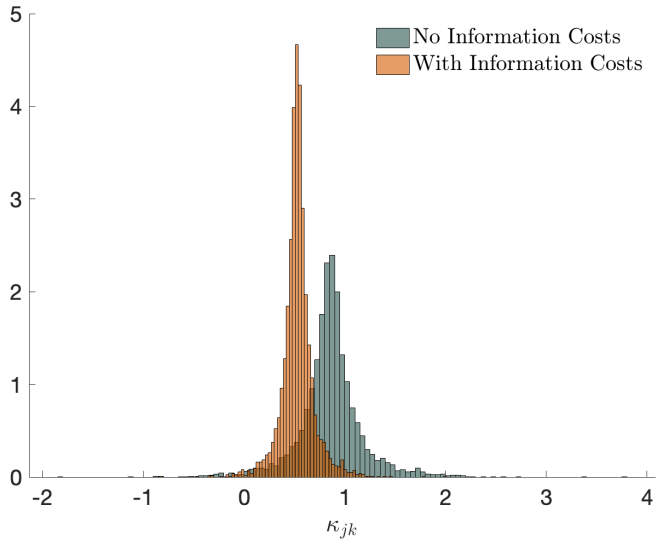


Figure 2.3: Estimated Migration Costs with and without Information Cost

The iterative estimation algorithm described in Section 2.4.2 can also be applied to this particular case. I obtain almost exactly the same estimates as in the direct estimation, which validates my iterative estimation method.

The distribution of bilateral migration costs resulting from the estimation of (2.26)—after projecting on origin and destination fixed effects to net out amenities—is displayed in Figure 2.3 for the period 2000–2007. Out of the $(J-1)^2 = 136^2 = 18,496$ migration trips that could be undertaken between any two mesoregions, only $10,382$ have positive flows recorded more than once during the first period. This implies that my procedure does not estimate a fixed bilateral resistance term for the pairs with no consistent flows. In figure 2.3, the average migration cost among the $10,382$ estimated bilateral costs is 0.94. For comparison, I report the distribution of the estimated migration costs for the same $10,382$ origin-destination pairs in the model with information frictions. With information frictions, the average migration cost in the first period is 0.56, so the average migration cost is 40% smaller once we allow for information frictions. This difference illustrates the quantitative relevance of the endogenous predispositions emanating from the information frictions.

To think of these costs as a fraction of income, consider what an agent earning some initial income w would have to receive as additional income Δw to be perfectly compensated for paying the average migration cost of 0.56. The additional income would require that $\log(w + \Delta w) = \log w - 0.56$, which corresponds to an increase $\Delta w/w$ of 75%, or an annualized flow of 3% of earnings, using a discount factor $\delta = 0.96$.

The estimated inverse elasticity from (2.26) in the first period is equal to $\nu = 2.92$, compared with 2.31 in the model with information frictions. The preference heterogeneity is then 21% smaller after accounting for information frictions. For the second period, I estimate $\nu = 3.12$, compared with 2.62, which corresponds to a 16% difference. All of these values are well in the range of the few existing estimates of the migration elasticity in the literature. For example, using a model similar to (2.21), Caliendo et al. (2018) estimate $\nu = 2.43$ in the context of migration between European countries during a similar period.

2.4.5 Migration Costs and Distance

I have estimated migration costs as origin-specific fixed effects. I now investigate how these migration costs vary with common measures of distance between regions. In particular, I express the migration costs κ_{jk} as a function of the euclidean distance between the population centroids of each region, as well as measures of bilateral travel times on the road network and a dummy for whether the two regions are contiguous. I rely on geo-referenced maps of the Brazilian road network from the Brazilian Ministry of Transportation for the year 2010 and compute travel time measures between each pair of regions using the Open Source Routing Machine. I estimate the following regression by OLS:

Table 2.3: Determinants of Migration Costs

	With Info Frictions		Without Info Frictions	
	2000-2007	2008-2014	2000-2007	2008-2014
Log Travel Time	0.384^a	0.370^a	0.619^a	0.568^a
	(0.053)	(0.051)	(0.089)	(0.077)
Log Distance	0.263^a	0.219^a	0.440^a	0.401^a
	(0.032)	(0.029)	(0.052)	(0.045)
Dummy Contiguous	-0.211^a	-0.182^a	-0.331^a	-0.301^a
	(0.012)	(0.013)	(0.021)	(0.016)
Observations	68,152	68,152	68,152	68,152
R^2	0.334	0.291	0.352	0.301

a indicates significant at the 1% level. In parenthesis, I report standard errors.

$$\kappa_{jk} = \beta_1 \log dist_{jk} + \beta_2 \log traveltime_{jk} + \beta_3 contiguous_{jk} + e_{jk}. \tag{2.27}$$

Table 2.3 reports the results from estimating (2.27) in each period. Migration costs are increasing with travel time and distance, and are smaller between contiguous regions. My estimates are slightly smaller in the second period, indicating migration costs have become slightly less dependent on distance and travel time. For comparison, the last two columns show the coefficients obtained after regressing the migration costs estimated from the model with no information frictions. The role of distance and travel time appear significantly more pronounced.

2.5 Testing Predictions of the Model

In this section, I present two exercises that illustrate the success of my model at describing migration patterns. In the first exercise, I uncover a set of new facts on migration patterns in Brazil. I show that the migration elasticity with respect to wages is decreasing with the distance between regions, increasing with the intensity

of past migration flows connecting them, and increasing with the internet penetration at the origin. I replicate the same empirical exercise in my model and show that the same patterns arise, with similar magnitudes, in the model with information frictions. In the model with complete information, the elasticities are constant. Second, the migration response to observed local positive shocks is more delayed for more distant origins and origins with lower internet penetration. Taken together, these two exercises support both the qualitative and quantitative relevance of the new mechanisms introduced in my model.

Before describing the two exercises, it is worth pointing out that logit models of migration with complete information such as (2.21) are typically considered quite successful at matching observed migration patterns. This is in large part thanks to their flexibility with respect to the bilateral migration costs κ_{jk}, which allow them to exactly match the average levels of bilateral migration flows between any two regions. Hence, in order to further test the ability of such models to accurately describe migration decisions, I focus on the cross-sectional variation in migration elasticities rather than levels, and on dynamic responses rather than time averages.

2.5.1 Heterogeneous Migration Elasticities

I start by presenting a simple empirical approach that provides sharply distinct predictions depending on whether information frictions are present or not. Note that in the model with complete information (2.21), there is a straightforward strategy for recovering the migration elasticity $1/\nu$, described in (2.26). I can however allow for some heterogeneity in the migration elasticity along some variable Z_{jkt}:

$$\Lambda_{jkt} = \beta_1 \Delta \log w_{jkt} + \beta_2 Z_{jkt} \times \Delta \log w_{jkt} + \beta_3 Z_{jkt} + D_{jk} + e_{jkt}. \tag{2.28}$$

For any variable Z_{jkt}, the model with complete information predicts that $\beta_2 = \beta_3 = 0$. However in the presence of information frictions, from equation (2.24), it is clear that the responsiveness of migration flows between two regions j and k with respect to productivity shocks θ_{kt} depends on the responsiveness of mean beliefs to θ_{kt}. If the beliefs held by agents in j about θ_{kt} increase when θ_{kt} increases, their migration elasticity is higher. These posterior beliefs are in turn determined by individual information acquisition and local information sharing.

In the presence of information frictions, the migration elasticity should decrease with distance. If, say, region k is close to region j, so that the migration cost κ_{jk} is small, then agents in j will pay quite a lot of attention to payoffs in k. Hence, upon receiving a recommendation to go to k, agents in j will update their beliefs significantly upward since they know this recommendation is likely to reflect the true productivity in k. This makes beliefs to close regions more responsive to nearby regions. In addition, when the productivity in this nearby region k is high at time $t-1$, people in k are more likely to stay in k and less likely to move to j. At the end of $t-1$ in j, there are fewer people from k, who tend to be pessimistic about k—people who leave their region tend to think their region is less attractive—and the shared beliefs in j about k become more optimistic. Since the productivity in k is persistent, it is likely to be high at time t as well, and agents in j are now likely to think productivity is high too. For these two reasons, beliefs about nearby regions are more responsive to productivity shocks, making the migration elasticity larger for nearby origin-destination pairs.

Second, the migration elasticity should also be increasing with the size of the past migration flows connecting an origin to another destination. In practice, if a large number of people in region j were in region k in the past, they are likely to be informed about this region and could pass along relevant information about the payoffs in k that will make migration more responsive to opportunities in k. In my

model, the effect of past flows from k to j on the responsiveness of contemporaneous migration from j to k is happening through local information sharing. If j tends to welcome a large number of people from k over time—maybe because the two regions are geographically close—beliefs in j about k tend to be accurate since they are largely influenced by the beliefs held by people coming from k. When a positive shock happens in region k, the flow from k to j will decrease as more people decide to stay. This translates into an increase in the shared mean beliefs about k in j due to the adjustment of population on the extensive margin.

Third, the migration elasticity should be increasing with local internet penetration. In practice, people with access to internet should have access to cheaper information and decide to gather more accurate information about migration opportunities. This should induce these people to take advantage of positive shocks and make them less likely to move when the conditions in the destination are not favorable. The idea that expanded access to information technology could explain changes in aggregate migration patterns was advanced by Kaplan and Schulhofer-Wohl (2017). In my model, regions with better access to internet also appear to have lower information costs λ_j. They are therefore able to form beliefs that are more accurate on average, so that in particular their beliefs about a region's payoffs are more likely to be high when the payoffs are high.

In Table 2.4, I report the results from estimating equation (2.28) over the period 2008–2014 for three different variables Z_{jkt}: the log distance between the origin and destination, the average fraction of residents with an active internet connection over the period, and the log number of individuals who moved from k to j in the past year. In column 1, I report the results from estimating equation (2.28) with observed migration flows and wages over the period 2008–2014. Column 2 contains the results from estimating equation (2.28) with migration flows and wages predicted by the model using the estimated parameters in the period 2008–2014. Once I solve for the

Table 2.4: Heterogeneity of Migration Elasticities in the Data and in the Model

		(1) Data	(2) Model With Info frictions	(3) Model Without Info frictions
A. Distance	Income gap	0.391^a	0.342^a	0.362^a
		(0.012)	(0.007)	(0.008)
	Log dist × Inc. gap	-0.012^a	-0.009^a	0.002
		(0.004)	(0.002)	(0.002)
B. Internet Access	Income gap	0.398^a	0.306^a	0.319^a
		(0.016)	(0.008)	(0.009)
	Internet × Inc. gap	0.132^a	0.101^a	0.003
		(0.015)	(0.004)	(0.004)
C. Past Mig. Flows	Income gap	0.292^a	0.241^a	0.311^a
		(0.018)	(0.008)	(0.008)
	Log Past Flows × Inc. gap	0.003^a	0.006^a	0.000
		(0.001)	(0.001)	(0.001)
	Observations	68,152	68,152	68,152

a indicates significant at the 1% level. In parentheses, I report two-way cluster robust standard errors that allow for correlation in regression residuals at the origin-year level and at the destination-year level. I include origin-destination and year fixed effects in each regression. The income gap is the difference in the log average wages between the destination and the origin. The effect of the level of Z_{jkt} is absorbed by the origin-destination fixed effect for both distance and average internet access, and is not reported for log past flows.

beliefs functions, value functions and mobility rule, I evaluate them at the observed L_{t-1} and recovered θ_t from the data. The interaction coefficients therefore reflect the differential correlation between the wage gap and the omitted expectational errors along the variable Z_{jkt} of interest. Column 3 reports the results from estimating (2.28) with migration flows and wages predicted by the model with no information frictions. As expected, the coefficients on the interactions from these regressions are all precisely estimated zeros.

In Panel A, column 1, the estimated coefficient of 0.391 for the income gap corresponds to the migration elasticity between adjacent pairs, while the negative coefficient for the interaction of the wage gap with log distance indicates that a 1 percentage point increase in distance is associated with a -0.012 decrease in the migration elasticity. For a pair of regions separated by the average distance of 500km, the estimated elasticity is therefore 0.3164. Note that this estimate is in line with the average inverse elasticity of $\nu = 3.12$ estimated in Section 2.4.4. There is, therefore, significant variation of the migration elasticity with distance.[38] In Panel A, column 2, the regression in the model provides strikingly similar results. Distance has a similar negative effect on the migration elasticity, slightly smaller but of the same order of magnitude. I cannot reject that the coefficients are the same at the 10% level. This similarity strongly emphasizes the quantitative relevance of information frictions, as it indicates that most of the decline in the responsiveness of migration flows with distance can be accounted for by the less effective acquisition and transmission of information between remote regions.

In Panel B, column 1, the coefficient of 0.281 represents the migration elasticity for individuals in regions with no internet penetration at all. The large coefficient of 0.132 indicates that the estimated elasticity is 0.413 for a region with complete internet penetration. In Panel B, column 2, I report the results from running the

[38]In their analysis of international migration flows, Bertoli et al. (2019) estimate a static gravity equation and find, similarly to my results, that the migration elasticity is decreasing with distance.

same regression in the model, also interacting the income gap with the measure of internet penetration at the origin. Here too, the model predicts a significant effect of having better internet access on the elasticity of migration. The positive coefficient reflects the strong negative correlation between internet access and λ_j documented in Section 2.4.3. The fact that the model can generate this positive interaction coefficient with internet access is another success, especially since the local information costs λ_j were estimated as fixed effects without imposing any relationship with respect to internet access. The model is able to account for 76% of the variation of the migration elasticity with local internet access observed in the data.

In Panel C, column 1, the coefficient of 0.362 corresponds to the estimated elasticity of migration if the log of the past migration flow was zero, namely if virtually all of the individuals in origin j were in k at the previous period. In practice, the fraction of current residents in a region who were in any other particular region k in the previous period is of the order of 10^{-6}, leading to an average migration elasticity of 0.319. In Panel C, column 2, I report the results from the same regression in the model. As expected, the migration elasticity with respect to wages in a destination is significantly larger when this destination is connected to the origin via large recent migration flows. The effects are about twice as large in the model as in the data, which is not surprising since the past migration flow is exactly the metric governing the adjustment of beliefs through local information sharing in the model. Taken together, these results show that introducing information frictions in the model delivers predictions that match important characteristics of migration flows.

2.5.2 Delayed Response to Local Shocks

I now confront the predictions of the model regarding the dynamic response of migration flows to local shocks. I show that the model is able to replicate a differential delay observed in the data in the migration response from some regions depending

on their distance to the shock and depending on their access to internet. In this section, I rely on an empirical approach developed by Fujiwara et al. (2020) to study the migration response to local shocks in Brazil.

In the model with perfect information (2.21), the delay with which individuals react to a positive local shock in another location is determined by the past migration flows between their origin and the location of the shock. If a large number of workers moved from the origin to the location of the shock in the period before the shock, this indicates that a significant number of agents would move to this location were the payoffs to increase. In this case, the migration response would be faster, with a large influx of migrants in the early periods after the shock is realized. In the model with information frictions, the delay in the migration response from a particular origin can vary even between regions that sent similar amounts of migrants to the location of the shock in the previous period. This is because people in two regions sending a similar fraction of migrants to a given destination in a period may have different information about the destination. For example, the past migration flow to the destination from some distant origin may be the same as from another nearby origin if, say, the distant origin experienced a negative shock so that many people decided to leave. In this case, agents from the distant region may not be as well informed about the destination as migrants from closer regions.

In order to describe the speed of the migration response of an origin region j to the destination k following a local shock in k at time t, I define the rate of migration from j to k at $t+s$ with respect to a horizon $(\underline{t}, \bar{t})$ as the fraction of total migrants who move from j to k at $t+s$, relative to the total number of migrants who will move from j to k between $t+\underline{t}$ and $t+\bar{t}$. Denoting by $L_{jkt+s} = p_{jkt+s}L_{jt+s-1}$ the gross migration flow between j and k at $t+s$, the rate of migration $p^{(\underline{t},\bar{t})}_{jkt+s}$ from j to k at $t+s$ with respect to a horizon $(\underline{t}, \bar{t})$ is

$$p_{jkt+s}^{(\underline{t},\bar{t})} = \frac{L_{jkt+s}}{\sum_{s'=\underline{t}}^{\bar{t}} L_{jkt+s'}}.$$

I also define the expected time of migration between j and k over a horizon $(\underline{t},\bar{t})$ as $\sum_{s=0}^{\bar{t}} s p_{jkt+s}^{(\underline{t},\bar{t})}$. I then say that region j has a faster migration response to k than j' at the horizon $(\underline{t},\bar{t})$ if the expected time of migration from j is smaller than from j'. According to this definition, a region j has a faster migration response to k than another region j' if the individuals who migrate from j to k do so earlier than migrants from j'.

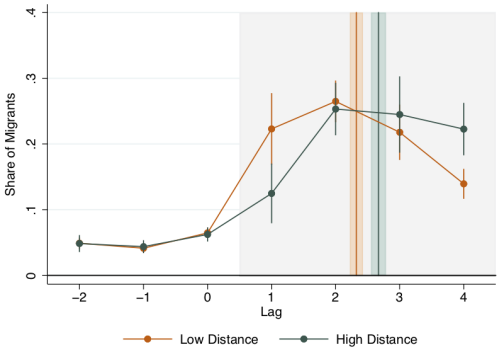
For the labor demand shocks listed in Appendix B.4.3, I consider migration flows to the shocked region in the two years prior or in the four years subsequent to the labor demand shock, so that $\underline{t} = -2$ and $\bar{t} = 4$, and estimate the following regression model:

$$p_{jkt+s}^{(\underline{t},\bar{t})} = \sum_{s'=-2}^{4} \mathbb{1}\{s = s'\} \left(\gamma_{1s'} int_{jt} + \gamma_{2s'} \log dist_{jk} + \gamma_{3s'} \log p_{jk}^{past}\right) + u_{jkt+s}, \quad (2.29)$$

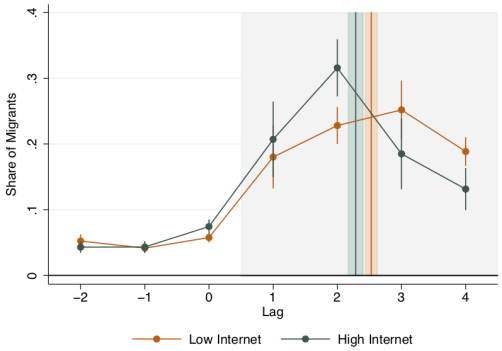
where t indicates the year in which the labor demand shock in the region of interest took place, int_{jt} is the internet penetration in region j at the year of the shock, $dist_{jk}$ is the distance between j and k, p_{jk}^{past} is the average migration probability from j to k in the three years preceding $t-2$, u_{jkt+s} captures unobserved determinants of the rate of migration at $t+s$, and $\{\gamma_{nt+s}; n = 1,2,3\}$ is the parameter vector of interest.

Figure B.4.2a and Figure B.4.2b display the estimates of $\{\gamma_{1t+s}, \gamma_{2t+s}\}$ that indicates how internet access, distance and past migration flows affect the timing of migration. I focus here on the largest labor demand shock in the sample, which took place in Ipojuca in the region of Recife in 2009 following the construction of a large refinery. The figures illustrate, for each year between $t-2$ and $t+4$, the predicted migration probability when the corresponding covariate, distance or internet access, is set to its 25th percentile, labeled as "Low", or to its 75th percentile, labeled as "High",

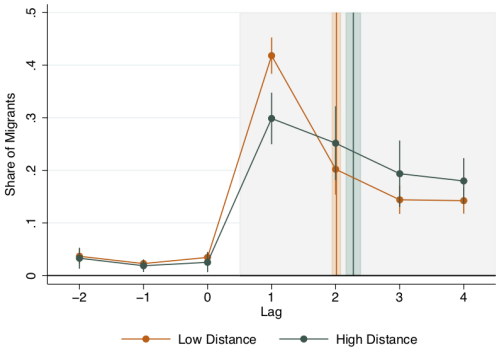
Figure 2.4: Delay in the Migration Response to the Local Shock in Ipojuca



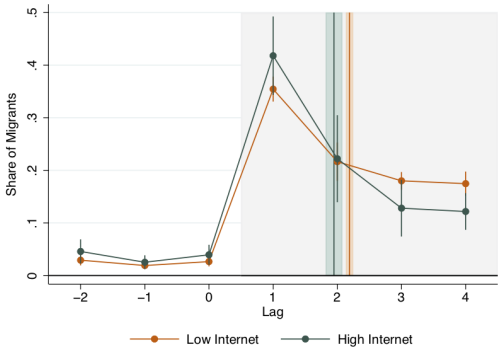
(a) Distance: Data



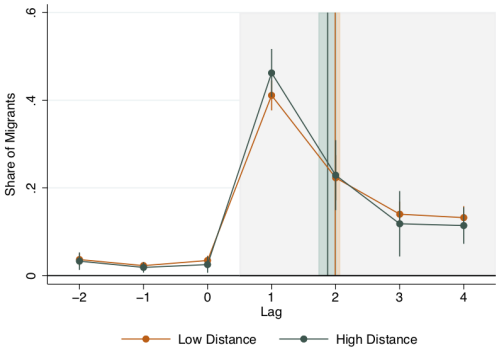
(b) Internet: Data



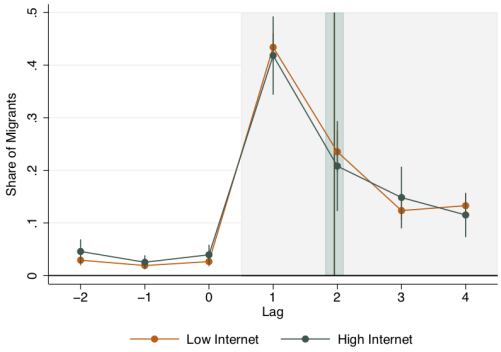
(c) Distance: Model with information frictions



(d) Internet: Model with information frictions



(e) Distance: Model without information frictions



(f) Internet: Model without information frictions

Note: I compute standard errors of the implied expected probability and the expected time of migration with the Delta method, and I cluster standard errors for estimates of the parameter $\{\gamma_{1t+s}, \gamma_{2t+s}\}$ by year and mesoregion of origin.

while other covariates are set to their mean values. The whiskers attached to each dot represent the 95% confidence interval for each predicted migration probability. The vertical lines indicate the expected time of migration between j and k once the shock is realized, and the shaded regions illustrate the corresponding 95% confidence intervals. Everything else equal, workers living in geographically close regions or in regions with higher broadband internet penetration react faster to the positive labor demand shocks happening in Ipojuca.

I then run regression (2.29) in the model, starting with the model with information frictions. Selecting the destination of interest k to be the mesoregion of Recife, I evaluate the migration flows predicted by the model under the observed population and recovered productivity vectors. The recovered productivity features a persistent increase starting in 2009 (see Appendix B.4.4), leading to an inflow of migrants similar to what is observed in the data. Although the model predicts that most of the migration should happen in the first period after the shock, Figure 2.4c illustrates that the estimated coefficients on distance lead to a significant delay for regions that are farther from to Recife. Figure 2.4d shows that the model also predicts a significant delay from regions with lower internet access.

In the model with no information frictions, Figures 2.4e and 2.4f reveal that the migration patterns resulting from the exact same simulated shock do not result in any significant delay along distance or internet access once we control for the past migration shares. This result confirms that the model with information frictions can generate migration patterns similar to those observed in the data, here the differential dynamic response to local shocks. In Appendix B.4.4, I show that this differential delay holds for other local shocks occurring in Brazil between 2000 and 2014.

2.6 Counterfactual Exercises

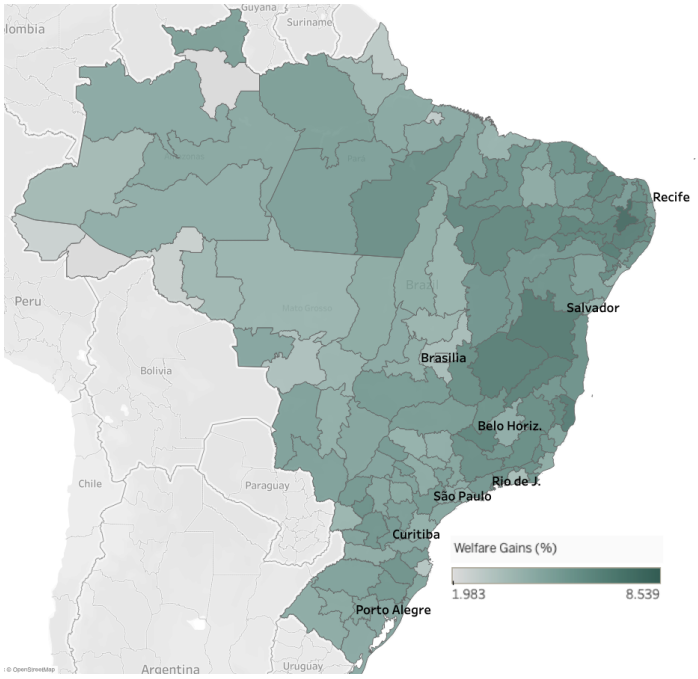
I have discussed several implications of information frictions on migration patterns. The main results of interest are that individuals feature endogenous predispositions towards moving to some regions, leading to lower estimated migration costs, and that their responsiveness to variations in payoffs is limited, leading to heterogeneous migration elasticities and delayed migration responses along geographic distance or internet access. I have verified these predictions in the data, and shown that the magnitude of the effects in the estimated model are close to the observed ones. To further illustrate the quantitative implications of my model for the spatial allocation of workers and welfare, I conduct a number of counterfactual exercises. First, I compare the outcomes in the economy estimated in the first period to a hypothetical one without information frictions. In a second exercise, I reduce information frictions by the amount corresponding to the estimated effect of increased internet access in the second period. For each of the exercises, I decompose the welfare gains into several contributions highlighting the role of the adjustment of the information structure for the overall effects.

2.6.1 Removing Information Costs

In this first counterfactual exercise, I evaluate the potential gains from completely removing information frictions. This exercise speaks to the initial question formulated in the introduction: what is the scope for improving the spatial allocation of workers by expanding their access to information, while taking into account that information must be acquired at a cost and that it can be shared locally?

To evaluate the effects of removing all information frictions, I first compute the expected value of being in a given region in the steady state of the first period, 2000–2007. In the stochastic steady state, the value of residing in a given region

Figure 2.5: Welfare Gains by Mesoregion after Setting All Information Costs to Zero



varies over time as different productivity shocks and population vectors are realized. Therefore, I report the expected value over possible realizations of (θ_t, L_{t-1}) in the ergodic distribution. Denoting by $\mathcal{V}_j$ the expected value of a region,

$$\mathcal{V}_j = \mathbb{E}_{\theta_t, L_{t-1}} \left[\sum_k \bar{p}_{jk}(\theta_t, L_{t-1}) \left(\bar{u}_{jk}(\theta_t, L_{t-1}) + \delta \bar{V}_k(\theta_t, L_{t-1}) \right) - I_j(\pi_j(L_{t-1})) \right]. \tag{2.30}$$

The expected value $\mathcal{V}_j$ can be computed by simulating the model over a large number of periods, so that the visited states $\{(\theta_t, L_{t-1})\}_t$ are representative of the ergodic distribution of possible states, and taking the simple average of $V_j(\theta_t, L_{t-1})$ over time. I then simulate a new steady state with all the parameters set to their value estimated in the first period, except for the information costs, which I set to zero. In this new stochastic steady state, I compute the expected values $\mathcal{V}'_j$ in each region.

Figure 2.5 displays the percentage change in the expected value of each region $\Delta\mathcal{V}_j/\mathcal{V}_j$ between the two equilibria, where I now define $\Delta X = X' - X$. The average welfare gain across regions is 5.55%. There is important heterogeneity across regions,

with gains ranging from only 2% in the most remote regions to 8.5% in a region close to Recife in the North-East. By looking at the initial distribution of information costs illustrated in Figure 2.2a, one might have expected the largest gains to accrue to regions starting with initially high information costs, such as the remote regions in the Amazon and regions of the North-East, and the lowest gains for regions with initially quite low information costs, such as Brasilia, São Paulo, and Rio de Janeiro. Figure 2.5 illustrates that this is mostly verified, although some regions in the Amazon do not seem to benefit as much as expected.

In the steady state with complete information, the standard deviation of earnings across space decreases by 15%. This is a clear illustration of the improved arbitrage of local shocks: a positive local shock attracts more immigrants in this economy, driving the local wage down. Similarly, when a negative productivity shock hits a region, more people leave to other regions offering better payoffs, thereby alleviating the negative wage effects in their origin region.

Perhaps surprisingly, the overall migration flows, computed as the fraction of the population moving to another region each period, decrease by 4% in the model with complete information. This net decline in migration flows masks two countervailing forces, both stemming from the fact that agents make fewer mistakes. First, and counteracting the decline in migration, agents with complete information no longer feature any predisposition towards staying in their current location. In the presence of positive costs of acquiring information about other regions, this predisposition was a reason for agents to stay in their current region. Second, agents now only move to regions that offer high payoffs, and no longer visit a region by mistake, so they move less often. The persistence in the productivity shocks allows workers to benefit from a mobility decision for several periods, reducing their propensity to move. This second effect appears to be dominating and leads to the 4% decline in overall geographic mobility. Note that this decline in migration flows corresponds to a comparison of the

two steady states, once each region's population has reached its ergodic distribution. One may expect that the transition from the first steady state to the second may lead to a temporary increase in mobility flows as workers relocate to the regions offering higher average payoffs.

To describe the forces at play even further, I decompose the welfare gains $\Delta\mathcal{V}_j$ into the contribution of three intuitive channels. Denote by $\bar{U}_j = (\bar{u}_{jk} + \delta\bar{V}_k)_{k=1,\dots,J}$ the vector composed of the sum of flow payoffs and future values for all possible destinations and states, (θ_t, L_{t-1}). Similarly, denote by $\bar{p}_j = (\bar{p}_{jk})_{k=1,\dots,J}$ the vector composed of the mobility probabilities to all destinations for all states. I can express the difference in expected value in a region j from (2.30) as

$$\Delta\mathcal{V}_j = \underbrace{\mathbb{E}_{\theta,L}\left[\Delta\bar{p}_j \cdot \bar{U}'_j\right]}_{\text{better sorting}} + \underbrace{\mathbb{E}_{\theta,L}\left[\bar{p}_j \cdot \Delta\bar{U}_j\right]}_{\text{better outcomes}} - \underbrace{\mathbb{E}_L\left[\Delta I_j\right]}_{\text{lower info cost}} . \tag{2.31}$$

The first term in the decomposition represents the welfare gains arising from better sorting of agents in the second equilibrium, relative to the initial equilibrium. It represents the expected gains in utility coming from the different mobility choices made by agents in the new equilibrium, maintaining the payoffs at their new value $\bar{U}'_j$. This better sorting can come about because agents have better information, leading them to chose the locations offering the highest payoffs. The second term reflects the gains due to the change in the payoffs themselves. The reallocation of population across regions can change the average wages offered in a region. Moreover, since the future value of residing in a region incorporates the expected payoff resulting from future mobility decisions, it can increase if the quality of the information available in the region has improved. The third term corresponds to the gains from spending less effort to acquire information. This decline can arise for two reasons. First, the cost of acquiring a given amount of information (λ_j) is cheaper. The second reason is due to the change in the precision of agents' prior beliefs. If the residents of the

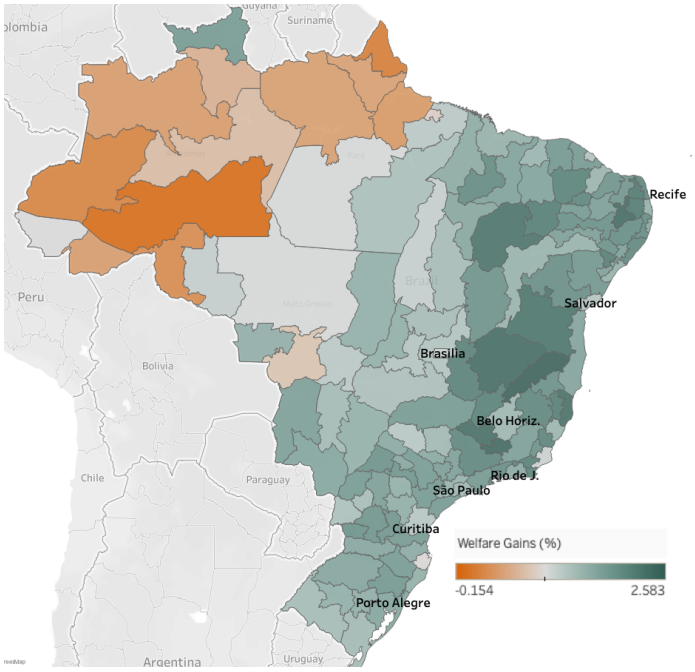
region have better information, local information sharing will allow agents to start with better information, so that they need to acquire less on their own. I find that 21% of the 5.5% welfare gains can be attributed to the "better sorting" channel, 58% to "better outcomes", and 21% to "lower information costs". It is remarkable that most of the gains seem to come from the response of mobility and earnings, so that the mechanical effect of lowering the information cost λ_j to zero accounts at most for about 20% of the gains.

2.6.2 Effect of the Expansion of Internet Access

I now evaluate a counterfactual decrease in information costs in each region by a magnitude equal to the estimated contribution of increases in local internet access. I use the measure of internet penetration observed in every region in 2014, the last year of the sample. Starting from the equilibrium in the first period, I decrease the information cost by an amount equal to the local internet access in 2014 (assuming that internet penetration was zero in the years 2000–2007), multiplied by the estimated coefficient ℓ_1 obtained after projecting the information costs on internet access. The average reduction in the information cost is about 0.73 units, starting from an initial average of 3.11.

Before describing the results, it is worth emphasizing that this exercise assumes that the only effect of increased local internet penetration is to allow workers to gather information more easily. It is likely that the expansion of internet access for households has been accompanied by a parallel expansion in access for firms, and that local firms' performance may have been altered by internet services. There is an extensive literature studying the effects of the development of information and communication technologies (ICT) on how production is organized. Hence, the results reported below should be interpreted as the effects of a hypothetical policy that would

Figure 2.6: Welfare Gains by Mesoregion after Expansion of Internet Access



reduce information costs by a magnitude similar to the result of expanded internet access, but without affecting the local productivity process.

Figure 2.6 depicts the geographic distribution of welfare gains from the counterfactual exercise. The average welfare gains amount to 1.63%. The standard deviation of earnings across space decreases by 4.01%, illustrating the better arbitrage of local shocks in the economy with lower information costs. The average decomposition of the welfare gains into the three channels described in Section 2.6.1 is almost exactly the same as in the previous exercise: better sorting, better outcomes and lower information costs account for 22%, 57% and 21% of the average welfare gains respectively.

In contrast to the previous counterfactual exercise, there are some regions that experience a mild decline in expected value in the equilibrium with lower information costs. This is the case for a few sparsely populated regions in the North-West of the country. Interestingly, these negative effects arise even though these regions have benefitted from increased internet access—although to a lesser extent than most other

regions. For instance, the region of Manaus experiences a decline in welfare of 0.15%, despite experiencing an increase in internet penetration of 23 percentage points.

To understand why some regions do not benefit from the episode of internet expansion, it is useful to decompose their welfare gains into the three channels described above. Focusing again on the example of Manaus, the contribution of information costs is actually negative, equal to -0.33. This means that agents in Manaus have to spend more to gather information than they did before. This happens because the information they can obtain from their local network has deteriorated. Indeed, in the new equilibrium, the population of Manaus has decreased by 7%, reflecting the fact that fewer workers from other regions now decide to locate there. Workers are now better informed and rarely find it optimal to move to the remote region of Manaus. In the initial equilibrium, Manaus would welcome workers with relatively good information who may have made a "mistake" by moving there. With fewer well informed visitors, local information sharing has become less effective in Manaus, and workers located there need to acquire more information on their own. Agents end up holding less precise information even after individual acquisition, and their migration decisions lead them to regions offering lower payoffs on average. This is reflected by a negative contribution of -0.36 of the "sorting" channel. Only the "outcomes" channel is working in favor of workers in Manaus, with a contribution of 0.54, as the lower population in the region tends to increase the average wage.

One additional reason why outcomes do not improve enough in Manaus to compensate for the decline due to worse information is that the information cost has decreased more in most other regions. As a result, when a positive local shock is realized, the people from better informed regions move faster to the location of the shock—as discussed in Section 2.5.2—and reap the benefits in the form of high wages before more people arrive and put downward pressure on wages.

2.7 Conclusion

In this paper, I propose a theory of migration under incomplete information. In my model, information about opportunities in other regions, earnings, and mobility patterns are all determined in equilibrium. I model, in a tractable way, both the incentives that agents have for acquiring information about some regions, and the possibility for workers to benefit from the information circulating in their local networks. When agents face costs of acquiring information, they collect limited information about regions offering a priori lower expected payoffs, making them less likely to move to these regions. Accounting for these endogenous default rules that hinder mobility appears to be quantitatively important for the estimation of bilateral migration costs.

I show that my model with information frictions can rationalize the observed heterogeneity in migration elasticities, as well as the differential delay in migration responses to local shocks from origins that are more distant or benefit from lower internet access. More generally, the model can help explain why the net inflow of migrants in response to positive local shocks often appears limited. Agents in regions where information about these shocks is more difficult to obtain, either because internet access is limited or because local networks can provide little relevant information, will be less likely to respond. I discuss how policies that could reduce information frictions can generate important welfare gains. Information acquired individually can be shared with other agents, inducing wide-ranging benefits from this positive externality. Yet, the distributional consequences from such policy interventions are not trivial. Some regions risk becoming "information traps" where agents struggle to gather accurate information.

The role of information frictions in migration decisions is likely to be more complex than the model I put forward in this paper. For example, local information sharing is likely to be more prominent between individuals of the same demographic

characteristics and working in similar occupations. Incorporating a richer description of local interactions could lead to interesting insights on migration decisions. Finally, studying an empirical setting featuring a clear distinction between payoffs observed and unobserved by workers could help further identify the contribution of information frictions.

Chapter 3

The Role of Establishment Size in the City-Size Earnings Premium in Spain

joint with Hannah Rubinton and Clara Santamaría

3.1 Introduction

Both large establishments and large cities are known to offer workers an earnings premium; facts referred to as the establishment-size and city-size earnings premium, respectively. We document a new fact: when workers move to larger cities, they move to larger establishments. Hence, part of the increase in earnings they experience can be accounted for by the fact that they transition to better-paying, larger establishments.

In this paper, we evaluate how much of the city-size earnings premium can be attributed to the establishment-size composition. Using rich administrative data on worker-level earnings histories in Spain, we find that on average 29.9 percent of the short-term earnings gain and 4.65 percent of the medium-term gain from moving to a larger city can be attributed to the transition to larger establishments.

Our results shed light on the drivers of the city-size earnings premium. The importance of establishment size suggests that explaining the presence of larger establishments in large cities is key for understanding the city-size earnings premium. However, we find that establishment size is not essential for the explaining the dynamic earnings gains of living in a big city, suggesting that the dynamic gains are likely to be driven by mechanisms that operate regardless of establishment size, such as learning or thicker matching markets.

We start by documenting three stylized facts. First, we document the well-known "city-size earnings premium". We find that moving to a city twice larger is associated with a 1.57 percent increase in earnings, in line with the findings of De la Roca and Puga (2017) using the same data.

Second, we establish what we call the "co-worker earnings premium",[1] namely that a transition to an establishment with more co-workers is associated with an increase in earnings. Controlling for observable and time invariant unobservable worker characteristics, we find that doubling the number of co-workers is associated with 1.90 percent higher earnings. We also document that the co-worker earnings premium is decreasing with city size. In Utrera, the smallest city in our sample, the co-worker earnings premium is more than twice as large as the premium in Madrid, the largest city.

Third, we document a new fact, the "city-size co-worker premium", corresponding to the increase in the number of co-workers associated with a move to a city twice larger. We find that the city-size co-worker premium is large; a worker moving to a city twice as large can expect 25.9 percent more co-workers[2].

[1]This is similar to what the literature has called the firm size or establishment size earnings premium. We use the terminology "co-worker earnings premium" because our estimates come from worker level data. Rather than estimating the correlation between pay per employee and firm size using firm or establishment level data as traditional in the literature on the establishment-size earnings premium, we estimate the correlation between an individual worker's pay and the size of the establishment at which they are employed, i.e. the number of co-workers.

[2]While Manning (2010) documents a positive correlation between establishment size and city size in establishment level data in the U.K. and the U.S., as far as we know, we are the first to use panel

We combine these facts to decompose the average city-size earnings premium into the contribution of increasing the number of co-workers, hence benefiting from the higher earnings offered by larger establishments, and a remaining "city-premium" not explained by the change in establishment size. Using the average co-worker earnings premium, we find that out of the total 1.57 percent increase in earnings from moving to a city twice as large, 39.5 percent is associated with workers moving to larger establishments.

Next, we move beyond the average city-size earnings premium to investigate the role of the establishment size in both the short-term and the medium-term city-size earnings premium. In order to do so, we include city-specific experience variables to control for experience accumulated in larger cities. Controlling for past experience, we find that moving to a city twice as large is associated with a short-term increase in earnings of 1.83 percent[3]. If we further control for the number of co-workers, we find that the increase in earnings falls to 1.28 percent. Therefore, the transition to larger establishments accounts for 29.9 percent of the short-term city-size earnings premium.

Finally, we explore the role of transitions between establishments of different sizes in explaining the dynamic earnings gains offered by large cities. These dynamic gains were first documented by De la Roca and Puga (2017), and reflect the additional earnings gains accruing to workers as they acquire additional years of experience in a large city. We find that after 8.9 years of experience in a city twice larger, which is

data to establish that workers moving to bigger cities tend to increase their number of co-workers. Further, we show this is true even when controlling for observable characteristics and worker fixed effects.

[3]Notice that the short-term city-size earnings premium is 0.2 percentage points larger than the average city-size earnings premium, despite earnings increasing faster in larger cities. This is due to a composition effect from experience. When we don't control for experience, part of the increase in earnings associated with moving to a larger city is due to the fact that moving to a larger city is correlated with being at the beginning of the career when the earnings growth is high. People moving to large cities tend to be younger and have little experience in large cities. On the contrary, people who move to smaller cities tend to have some experience in larger cities and their earnings fall a bit less.

the average experience in our sample, workers benefit from an additional 1.29 percent increase in earnings, in line with the findings in De la Roca and Puga (2017).

These dynamic gains could be due to a faster transition to larger establishments in large cities or a steeper gradient of earnings with establishment size in large cities – i.e., a larger co-worker earnings premium in large cities. Indeed, we document a slightly faster transition to larger establishments in large cities. However, as reported above, we find that the co-worker earnings premium is actually smaller in large cities. This fact dampens the contribution of the faster growth in the number of co-workers leading to a small dynamic gain from establishment size. Only 4.65 percent of the dynamic earnings gains in larger cities can be attributed to a faster growth in the number of co-workers.

This paper contributes to several strands of the existing literature. Our primary aim is to gauge the relative importance of establishment size and city-effects in the city-size earnings premium. The city-size wage premium and the establishment-size wage premium are typically associated with different mechanisms, and largely separate strands of literature. On the one hand, this paper is related to the recent literature employing rich micro-level data to understand the role of firm and establishment size in determining earnings (Bloom et al. (2019); Mueller et al. (2017))[4]. On the other hand, the paper echoes a vast literature exploring the sources of the city-size earnings premium including sorting of more productive workers to large cities, local agglomeration forces, knowledge spillovers, labor pooling and better worker-firm matching (Eeckhout et al. (2014); Diamond (2016b); Behrens et al. (2014); Giannone (2018); De la Roca et al. (2014); Combes et al. (2008); Davis and Dingel (2014); Dauth et al. (2019)). In principle, given the positive correlation between city size and establishment size, either of the two size premia could be driving the other one. Our

[4]This, in turn, echoes an older literature on the firm size earning premium including papers by Burdett and Mortensen (1998); Mellow (1982); Brown and Medoff (1989); Bayard and Troske (1999).

contribution is to quantify their relative importance and provide new evidence to be contrasted against some of the mechanisms proposed in the literature.

Our empirical findings contribute to the literature on firm sorting across city sizes which requires some complementarity between establishment and city size (Gaubert (2018); Duranton and Puga (2001); Brinkman et al. (2012); Combes et al. (2012); Manning (2010)). The robustness of the city-size co-worker premium provides evidence for this complementarity. Further, the fact that the co-worker earnings premium is smaller in large cities is consistent with the mechanism in Manning (2010), who argues that firms sort based on differences in monopsony power across cities. With lower monopsony power firms face a more elastic labor supply curve and do not need to raise wages as much to attract workers. This generates a weaker relationship between earnings and establishment size in large cities.

Finally, our results also connect to several theories advanced to explain the dynamic city-size earnings premium. In particular, the limited role of establishment size transitions in the dynamic gains favors stories of learning as in Davis and Dingel (2012). Finally, Eckert et al. (2019) find that part of the earnings premium for experience accumulated in large cities is associated to experience in high-paying establishments using quasi-experimental data for refugees in Denmark. In our data set, although we observe the whole history of locations, we do not observe the history of the number of co-workers. Instead, we focus on quantifying the dynamic effect coming from a faster growth in the number of co-workers in larger cities.

The rest of the paper is organized as follows. Section 3.2 describes the Spanish administrative data. Section 3.3 documents the new stylized facts on establishment size earnings premium, the establishment size earnings premium across cities, the city-size earnings premium, and the city-size co-worker premium. Section 3.4 presents the main results decomposing the city size earnings premium into the city premium and the co-worker premium and compares the contribution of establishment-size compo-

sition for short-term and medium-term gains of working in larger cities. Section 3.5 concludes.

3.2 Data Description

This paper uses the Muestra Continua de Vidas Laborales (MCVL). This dataset is maintained by the Social Security Administration (SSA) in Spain and it consists of an annual panel from 2006 to 2013. The MCVL sample is selected in two steps. In a first step, an algorithm is designed to select 4 percent of all personal ID numbers in use[5]. The number of individuals in this first selection is expanded by including a 4 percent of newly issued numbers every year[6]. In a second step, the selection is refined to include, from among those who are part of the first selection, the individuals who belong to the population of reference during the sample year. The population of reference is all the individuals that had a relationship with the SSA, either because they contributed with their labor earnings or because they received unemployment or pension benefits. These two-step procedure results in a 4 percent random non-stratified sample of the population of reference in every year. Moreover, the procedure guarantees a panel dimension that follows people across years even as they exit and enter the population of reference.

The data covers the years from 2006 to 2013. This period includes the 2008 financial crisis in Spain, as well as the housing boom leading up to it. The unemployment rate in Spain reached a historical low of 8 percent in 2007 and peaked in 2013, at approximately 26 percent. Monthly fixed effects are included in the estimation to control for the national trend.

[5]The personal ID number refers to the Documento Nacional de Identidad (DNI) number for nationals, and the Número de Identificación del extranjero (NIE) for foreign nationals. Every resident in Spain is required by law to acquire a DNI by the age of 14.

[6]The criteria for a number to be selected is based in a deterministic fashion on characteristics the number itself, and as such does not change across years an is orthogonal to any characteristic of the holder.

In the final sample we include employed working-age males born in Spain.[7] We exclude public employees, workers under apprenticeship contracts, co-op workers, and employees in agriculture, fishing, forestry, extraction industries, public administration, education, health services and international organizations. We also exclude workers whose contracts specify discontinuous involvement, and individuals who worked for less than the equivalent of 30 days in a calendar year. Finally, the sample does not include individuals working in the Basque Country or Navarre, since these regions keep their own fiscal records which are not managed by the Social Security Administration. This selection results in 7,019,750 monthly-person observations.[8] Section C.1.2 in the Appendix includes summary statistics of the final sample in Table C.1.1 and a discussion on how representative it is of the overall Spanish population along different dimensions. Moreover, we relegate to Section C.1.1 in the Appendix the precise definitions of urban area, establishment, employer, experience, and earnings.

One of the advantages of the data is that it includes the complete history of an individual relationship with the SSA since they first entered the labor force. This long-run dimension allows us to construct very detailed information on the work experience of individuals, including the number of years they spent in large cities. Being able to observe their whole labor history is key for our analysis.

[7] The reason to not include women is that female labor force participation in Spain was extremely low during the 80s and it would significantly affect our measure of labor experience in large cities for the older women in the sample.

[8] The sample spans 8 years, that is, 96 months. This means that if we observed everyone for the full time span, it would correspond to around 73 thousand persons, which is 4 percent of approximately 1.8 million people. In comparison, the population of employed males between 15 and 75 with Spanish citizenship is 8.3 million. This means that the additional selections on geography, industry, type of employment, and attachment to the labor market cut the population of reference to less than a quarter. There is a trade-off between having a sample of people whose earnings we can expect to behave according to the model, and having a non-representative sample of the Spanish population as a whole. We opted for the first one, but a word of caution is in place when extrapolating the results to the Spanish population, or populations who are very different from the population of reference.

3.3 Establishment-Size Composition and the City-Size Earnings Premium

In this section, we document the three main stylized facts. First, we document the city-size earnings premium, namely the positive association between earnings and city size. Second, we document the co-worker earnings premium, corresponding to the positive association between earnings and the number of co-workers. We further show that the co-worker earnings premium is not constant across cities. Within Utrera, the smallest city in Spain, moving to an establishment twice as large is associated with a 3.7% earnings premium, while in Madrid a worker who moves to an establishment twice as large can only expect a premium of 1.5%. Finally, we document the city-size co-worker premium, that is, the increase in the number of co-workers associated with living in a larger city.

3.3.1 Two step methodology

Before documenting the stylized facts, we briefly discuss the methodology. We follow De la Roca and Puga (2017) and employ a two-step procedure. In the first step, we regress the variable of interest on city fixed effects and controls. Second, we regress the city fixed effects on city size.

There are two reasons for employing the two step procedure. First, by using the city fixed effects to estimate the earnings and co-worker premium, we are giving equal weight to cities of different sizes. If instead we ran one regression of individual log earnings on city size, we would over-weight large cities that are better represented in the sample. Second, by doing the estimation in two steps, the city fixed effects absorb all the city-level shocks that would otherwise introduce a correlation structure in the error terms. Combes and Gobillon (2015) provide a detailed discussion of the

relationship between an equivalent one-step, in which city size is directly introduced as a co-variate, and the two-step procedure.

To fix notation, first we run the following regression of log earnings w_{it} (or log number of co-workers in some of the specifications) for individual i, at time t:

$$\ln(w_{it}) = \alpha_{c(i,t)} + \delta_t + \mu_i + x'_{it}\psi + \varepsilon_{it}, \tag{3.1}$$

where c(i,t) is the city where individual i lives at time t, $\alpha_{c(i,t)}$ is the corresponding city fixed effect, δ_t is a month-year fixed effect, μ_i is an individual fixed effect, x_{it} is a vector of characteristics of worker i at time t which include education, occupation, sector, experience, tenure, and type of contract; and ε_{it} is the residual earnings. For reference, section C.1.2 of the Appendix includes a table with the summary statistics for the regressors and variables of interest.

In the second step, we regress the estimated city fixed effects on the log of our measure of city size, that is, the population within 10 km of the average person:

$$\alpha_c = \beta \ln(city\, size_c) + u_c. \tag{3.2}$$

The city size earnings (or co-workers) premium is then defined as the elasticity of the city fixed effect, that is, the city-level average earnings (or co-workers) with respect to city size. The city size premium will be reported as the increase in earnings (or co-workers) associated with a 100 percent increase in city size. By estimating the city size premium in this way, it is explicit that we treat each city as one observation.

3.3.2 Three stylized facts

Table 3.1 presents the three main stylized facts. Panel A shows the city-size earnings premium, or the positive correlation between earnings and city size. The elasticity between earnings and city size is 0.0712 meaning that workers in cities with twice the

Table 3.1: City-Size Earnings Premium, Co-Worker Earnings Premium and City-Size coworker premium

	(1)	(2)	(3)
Panel A: City-Size Earnings Premium, Second Step			
	Ln Earnings City FE from 1st Step		
Ln City Size	0.0712	0.0469	0.0225
	(0.0107)***	(0.0082)***	(0.0051)***
Panel B: Co-Worker Earnings Premium			
	Ln Earnings		
Ln Co-workers	0.0672	0.0487	0.0271
	(0.0006)***	(0.0005)***	(0.0005)***
Panel C: City-Size Co-Worker Premium, Second Step			
	Ln Co-Workers City FE from 1st Step		
Ln City Size	0.3370	0.3207	0.3317
	(0.0493)***	(0.0480)***	(0.0465)***
Month FE	Y	Y	Y
Sector FE	Y	Y	Y
Worker FE	N	N	Y
Controls	N	Y	Y
Observations	7,019,750	7,019,750	7,019,750

Panel A reports the results of the second step, consisting in regressing the city fixed effects on log city size. In the first step, we regressed log earnings on city fixed effects and a series of controls. City size is measured as the number of people living within 10km of the average person. Panel B reports the results from regressing log earnings on log establishment workers. An establishment corresponds to a firm-province tax identifier. Panel C reports the results of the second step, consisting in regressing the city fixed effects on log city size. In the first step, we regressed log number establishment workers on city fixed effects and a series of controls. In all three panels, controls include education, occupational categories, industry and time fixed effects, and type of contract. In addition, we control always for month and sector fixed effects. Clustering is implemented at the person level. Information on worker fixed effects, controls and the number of observations refers to the first step of Panel A and C, and to Panel B. Data Source: Spain's Continuous Sample of Employment Histories 2006-2013.

population have on average 5.0 percent higher earnings[9]. A significant part of this difference can be accounted for by observable characteristics such as sector, occupation, education, experience, tenure, and whether the contract is part time. Once we control for worker characteristics, in column 2, the elasticity of earnings with respect to city size decreases to 0.0469, that is, more than a third of the percentage difference in earnings can be explained by observable characteristics.

Finally, in column 3, we control for unobserved worker characteristics by including worker fixed effects. The elasticity between earnings and city size decreases to 0.0225 percent meaning that a worker moving to a city with twice the size should expect an average 1.57 percent increase in earnings. The earnings premium is about half of the city-size earnings premium before including worker fixed effects.

There are two differences between the regressions with and without worker fixed effects that could drive this decrease. First, as intended, the decrease in the observed city-size earnings premium could be due to controlling for sorting in unobserved worker characteristics. Second, the decrease could be due to the change in the way the city fixed effects are identified once worker fixed effects are included. Without worker fixed effects, the city fixed effects reflect the difference in average residual earnings across cities. With worker fixed effects, the city fixed effects are the difference in average residual earnings for workers who ever move across cities. Therefore, once we control for worker fixed effects, we are identifying the city-size earnings premium using only workers who move across cities, these movers might be systematically different from workers who never move. Subsection C.2.1 of Appendix C.2 explores which factor is more important and finds that the city-size wage premium is not substantially different for the population of workers that ever move and that workers who move to a big city are not systematically moving to big establishments. These results suggest that our findings are not driven by selection.

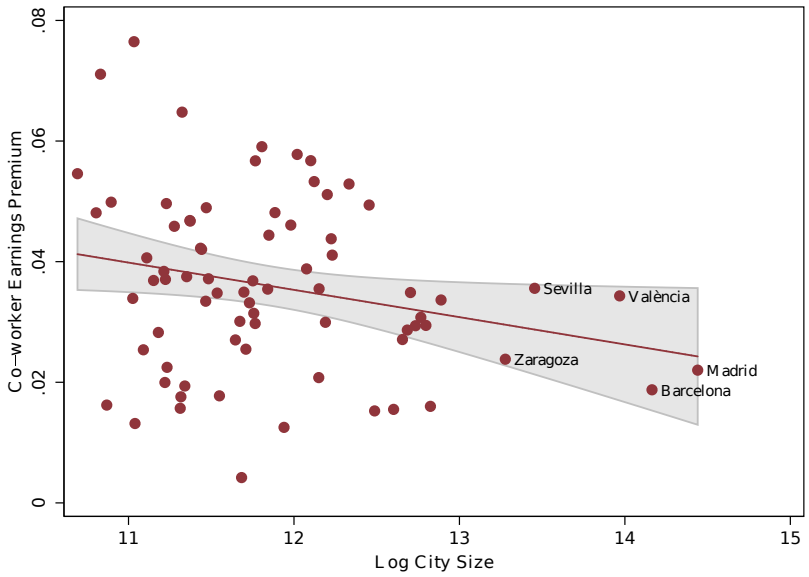
[9]The increase in earnings will be equal to $y'/y = (exp(.071 * \ln(x'/x)) - 1)$. If x' is twice the size of city x, then x'/x will be 2.

Next, Panel B of Table 3.1, documents the correlation between the number of co-workers and earnings in the data. The elasticity between earnings and the number of co-workers is 0.0672 meaning that workers with twice as many co-workers have, on average, 4.7 percent higher earnings. As with the city-size earnings premium, part of this effect could be due to workers sorting across establishments. Controlling for a broad set of observable worker characteristics reduces the elasticity to 0.0487 percent. This decrease is evidence that there is sorting of higher earners to establishments with more co-workers. In column 3, we further control for worker fixed effects and evaluate the relevance of unobserved characteristics. The coefficient on log co-workers captures the elasticity between earnings and establishment size for workers who move across establishment sizes. The elasticity decreases to 0.0271 percent meaning that a worker moving to an establishment with twice as many co-workers should expect a 1.90 percent increase in their earnings. We refer to this increase as the co-worker earnings premium.

It is worth noting that the city size earnings premium (2.2 percent) is close in magnitude to the co-worker earnings premium (2.7 percent). Implying that a worker switching to an employer with twice as many co-workers should expect a similar increase in earnings as if they were moving to a city with twice the size.

Moving to an establishment with more co-workers is associated with large earnings gains. However, it is possible that these gains are not the same across all cities. We calculate the within-city co-worker earnings premium, namely the expected premium from increasing the number of co-workers conditional on staying in the same city, and we show that it is decreasing in city size. Figure 3.1 shows the relationship between the co-worker earnings premium and city size. The within-city co-worker earnings premium is 0.37 percentage points smaller in cities twice as large. In practice, this corresponds to a significant difference in the co-worker earnings premium between large and small cities. In Utrera, the smallest city in our sample, the elasticity

Figure 3.1: Co-worker Earnings Premium by City



Note: This figure plots the coefficient on log number of coworkers on log city size, where the city specific coefficient results from regressing log earnings on sector, time and city fixed effects interacted with log number of workers,. City size is the average population within 10 km of the average person. The gray area corresponds to the 95% confidence interval. Data Source: Spain's Continuous Sample of Employment Histories 2006-2013.

between earnings and city size is .053 percent, while in Madrid the elasticity is only .022 percent.

This novel fact has important implications. In decomposing the city-size earnings premium, our goal is to understand what share of the city-size premium is driven by a contemporaneous move to a larger establishment. The fact that the within-city co-worker earnings premium is smaller in large cities means that the expected effect of moving from a small establishment in a small city to a large establishment in a large city will be smaller than the naive estimate of combining the average co-worker earnings premium and the average city-size co-worker premium. We provide more detail on how we calculate the within-city co-worker premium, show the complete set of results with and without worker fixed effects, and further discuss the implications of this novel finding in subsection C.2.2 of Appendix C.2.

Finally, Panel C of Table 3.1 documents the positive correlation between number of co-worker and city size. The elasticity between the number of co-workers and city size is 0.3370, meaning that workers in cities with twice the size have on average 26.3 percent more co-workers. Once we control for worker characteristics in column 2, the elasticity only decreases to 0.3207. Finally, controlling for worker fixed effects in column 3 only decreases the elasticity to 0.3317, meaning that a worker moving to a city with twice the size should expect an average 25.9 percent increase in the number of co-workers. The estimates in the three specifications are very close, which suggests that sorting does not play an important role in explaining the city-size co-worker premium. That is, workers who select into larger cities tend to have characteristics that make them higher earners, as shown in Table 3.1, but these characteristics do not make them significantly more likely to work with more co-workers.

It may not come as a surprise that local employment is larger in large cities if firm size is driven by demand. This is what we would expect in service or non-tradable sectors where employment and demand may be tightly linked to the local market size. In order to test whether non-tradable sectors are driving most of the differences in establishment size, we follow our two-step methodology described in section 3.3.1 and, for each sector, regress the number of establishment workers on characteristics of the workers and city fixed effects. Then we regress the city fixed effects against log city size. Table 3.2 presents the elasticity of co-workers with respect to city size for different sectors. The larger elasticities are found in finance, communication, research and development, professional services and computer sectors, all highly tradable sectors while the least tradable sectors such as personal services and hotels present the lowest elasticities. Thus, we conclude that the elasticity of the number of co-workers with respect to city size is not driven by local market size.

Table 3.2: City Size Co-worker Premium by Sector

Tradables		Non-tradables	
Manufacturing	0.1474	Real Estate	-0.4101
	(0.1462)		(0.1433)***
Communication	0.5407	Rental Services	0.6483
	(0.1375)***		(0.3110)**
Professional Services	0.5979	Construction	0.2019
	(0.1257)***		(0.0659)***
Computer	0.7517	Retail and Wholesale	0.3355
	(0.1264)***		(0.0982)***
Research and Development	0.5074	Hotels	0.3551
	(0.1345)***		(0.1045)***
Finance	0.6513	Transportation	0.1041
	(0.1614)***		(0.1385)
		Entertainment	0.2281
			(0.1102)**
		Personal Services	-0.1431
			(0.1414)

Note: * p<0.1; ** p<0.05, ***p<0.01. This table reports the elasticity with respect to city size of a city fixed effect resulting from regressing log establishment workers on city fixed effects sector by sector, controlling for experience, tenure, occupation, type of contract, time fixed effects, and individual fixed effects. City size is the number of people working within 10km of the average person and an establishment corresponds to a firm-province tax identifier. Clustering is implemented at the person level. Data Source: Spain's Continuous Sample of Employment Histories 2006-2013.

3.4 The Role of Establishment Size in the City-Size Earnings Premium

In this section, we explore the role of the city-size co-worker premium, i.e. the fact that workers in larger cities tend to have more co-workers, in generating the city-size earnings premium. The goal is to compare earnings of similar workers with a similar number of co-workers in order to measure the effects of city size that do not work through differences in the number of co-workers. If a large fraction of the city-size earnings premium disappears when comparing workers with a similar number of co-workers, this would imply that in order to understand the city-size earnings premium,

one must understand why the number of co-workers is higher in larger cities. If on the contrary, the earnings differences remain when comparing workers in similarly sized establishments, this would highlight the importance of city size effects that do not act through the number of co-workers.

3.4.1 The Role of Establishment Size in the Average City-Size Earnings Premium

We start by documenting how much of the city-size earnings premium can be attributed to the city-size co-worker premium, that is the fact that workers in large cities have, on average, more co-workers. We then document how much of the co-worker earnings premium is due to the fact that workers in large cities have more co-workers. In order to answer these questions, we use the two-step methodology described in the previous section. In the first stage, we regress log earnings on city size fixed effects and workers characteristics, this time controlling for the number of co-workers. In the second stage, we regress the city fixed effects on log city size. We obtain a new city-size earnings premium corresponding to the increase in earnings associated to a transition to a larger city while leaving the number of co-workers unchanged. The results for this exercise are included in Table 3.3.

First, we analyze the change in the city-size earnings premium that results from controlling for the number of co-workers. Column 1 of Table 3.3 reports an elasticity of earnings with respect to city size of 0.0469. This indicates that workers with similar characteristics, not necessarily working with the same number of co-workers, but living in a city with twice the size have 3.45 percent higher earnings.[10] In column 2 of Table 3.3, the elasticity has decreased to 0.0329, indicating that workers with similar characteristics and working with the same number of co-workers but living

[10]Column 1 of Table 3.3 reproduces the city-size earnings premium obtained in column 2, Panel A of Table 3.1.

Table 3.3: The Role of Transitions to Larger Establishments in the Average City-Size Earnings Premium

	(1)	(2)	(3)	(4)
	Ln Earnings City FE from Step 1			
Ln City Size	0.0469	0.0329	0.0225	0.0136
	(0.0082)***	(0.0073)***	(0.0051)***	(0.0048)***
Observations	76	76	76	76
R^2	0.234	0.152	0.158	0.073
	Step 1: Ln Earnings			
Ln Co-Workers		0.0437		0.0269
		(0.0005)***		(0.0005)***
City FE	Y	Y	Y	Y
Worker FE	N	N	Y	Y
Controls	Y	Y	Y	Y
Observations	7,019,750	7,019,750	7,019,750	7,019,750
R^2	0.491	0.514	0.462	0.454

This table reports the results from regressing individual log earnings on city fixed effects and a series of controls, and then regressing the city fixed effects on log city size. City size is the number of people living within 10km of the average person. Controls include education, occupational categories, industry and time fixed effects, and type of contract. Clustering is implemented at the person level. Data Source: Spain's Continuous Sample of Employment Histories 2006-2013.

in a city with twice the size have 2.31 percent higher earnings. The 1.14 percentage points decrease in the cross-sectional comparison of earnings may be due to both the fact that by comparing workers with similar number of co-workers, we may be comparing more similar workers in terms of unobserved characteristics, and that, even for a particular worker, the earnings premium of moving to a larger city may be in part due to an increase in the number co-workers.

In order to disentangle the two possible sources of difference in the elasticities in columns 1 and 2, we account for differences in the selection on time invariant unobserved characteristics by including worker fixed effects in the first stage. In

column 3, we obtain an elasticity of 0.025. This indicates that a worker moving to a city with twice the size and not necessarily staying with the same number of co-workers should expect a 1.57 percent increase in earnings on average.[11] In comparison, column 4 reports an elasticity of 0.0136. This implies that a worker moving to a city with twice the size but staying with the same number of co-workers should expect a 0.95 percent increase in earnings. The 0.62 percentage points decrease implies that 39.5 percent of a worker's expected increase in earnings in a city with twice the size is associated to the increase in the number of co-workers.

Second, we investigate the change in the co-workers premium as a result of including city fixed effects. The coefficient in the first step of 0.0437, reported in column 2 of 3.3, shows that workers with similar characteristics and living in the same city but with twice as many co-workers have 3.08 percent higher earnings. In comparison, in column 2, Panel B of Table 3.1, the elasticity was 0.0487, indicating that workers with twice as many co-workers but not necessarily working in the same city have 3.43 percent higher earnings. Hence, the fact that by comparing workers with different number of co-workers, we may be comparing workers in different cities, can account for at most 10.2 percent of the earnings difference.

Finally, we control for unobserved characteristics by including worker fixed effects in Column 4 of Table 3.3. We find an elasticity of 0.0269. A worker moving to an employer with twice as many co-workers but staying in the same city, should expect 1.88 percent higher earnings, the same earnings increase as a worker moving to an employer with twice as many co-workers but not necessarily staying in the same city.[12] Therefore, the co-worker earnings premium cannot be explained by workers that increase their number of co-workers by simultaneously moving to a larger city.

[11] Column 3 of Table 3.3 reproduces column 3, Panel A of Table 3.1.
[12] As reported in column 3, Panel B of Table 3.1.

3.4.2 Decomposing the Short-Term and Medium-Term Effects of Establishment Size

Until now, we have focused on the effect of increasing either city size or the number of co-workers on a workers' average earnings, since we were not controlling for the fact that some workers had accumulated experience in large cities and others in small cities. Previous work by De la Roca and Puga (2017) documented that experience accumulated in larger cities results in faster earnings growth, and that this accounted for a significant fraction of average city-size earnings premium. In this section, we look at the effect of controlling for the number of co-workers on the estimation of both the short-term effect, and the effect in the medium-term separately. We test whether accumulating experience in larger cities may increase earnings in part by allowing workers to move to establishments with more co-workers. If this is the case, controlling for the number of co-workers would reduce the positive effect of large-city experience on earnings.

In order to study dynamic effects, we follow the two-step procedure used by De la Roca and Puga (2017). In a first step, we regress log earnings (or log number of co-workers) on city fixed effects and the number of years of experience working in each of the cities. Let's assume that wages for individual i, at time t are given by:

$$\ln\left(w_{it}\right) = \alpha_{c(i,t)} + \sum_{j=1}^{C} \delta_{jc(i,t)} e_{ijt} + \mu_i + x'_{it}\beta + \varepsilon_{it}, \tag{3.3}$$

where $c\left(i,t\right)$ is the city where individual i lives at time t, $\alpha_{c(i,t)}$ is a city fixed effect, δ_{jc} represents the value of experience acquired in city j when the worker is currently in city c, e_{ijt} is the number of years individual i worked in city j up to time t, μ_i is an individual fixed effect, x_{it} is a vector of characteristics of worker i at time t which include education, occupation, sector, experience, tenure, and type of contract; and ε_{it} is the residual wage.

We define the short-term city-size premium β^{st} as the coefficient on log size from the following regression:

$$\alpha_c = \beta^{st} \ln (city\ size_c) + u_c. \tag{3.4}$$

The medium-term city-size premium β^{mt} is defined as the coefficient on log size from the following regression:

$$\alpha_c + \delta_{cc} mean\ experience = \beta^{mt} \ln (city\ size_c) + u_c, \tag{3.5}$$

so the medium-term city-size premium captures both the short-term earnings gains and the specific speed of earnings growth with experience associated to living in a larger city.

The results are shown in Table 3.4. For the reasons discussed in Section 3.4.1, we include worker fixed effects in all the regressions. First, we decompose the short-term city-size earnings premium into the part that is associated with the increase in the number of co-workers, and the part that is independent of the number of co-workers. In column 1, the short-term city-size earnings premium is 0.0261. This implies that a worker moving to a city with twice the size should expect a 1.83 percent short-term increase in earnings. This increase includes the effect of the increase in the number of co-workers. In column 2, after controlling for the number of co-workers, the short-term city-size earnings premium decreases to 0.0183. This corresponds to a 1.28 percent increase in earnings from moving to a city with twice the size. Therefore, the increase in the number of co-workers can account for 29.9 percent of the short-term city-size earnings premium.

Next, we decompose the medium-term city-size earnings premium. In column 3, the medium-term city-size earnings premium is 0.043. Hence, a worker moving to a city with twice the size should expect a 3.12 percent increase after 8.9 years. Thus,

Table 3.4: The Role of Transitions to Larger Establishments for Short-Term and Medium Term City-Size Earnings Premium

	(1)	(2)	(3)	(4)
	Short-Term: Ln Earnings City FE from Step 1		Medium-Term: Ln Earnings City FE Plus City-Size Specific Exper. from Step 1	
Ln City Size	0.026	0.018	0.044	0.035
	(0.0050)***	(0.0048)***	(0.0100)***	(0.0093)***
Observations	76	76	76	76
R^2	0.201	0.126	0.358	0.296
	Step 1: Ln Earnings			
Ln Co-Workers		0.027		0.027
		(0.0005)***		(0.0005)***
City FE	Y	Y	Y	Y
Worker FE	Y	Y	Y	Y
City-Specific Exper.	Y	Y	Y	Y
Controls	Y	Y	Y	Y
Observations	7,019,750	7,019,750	7,019,750	7,019,750
R^2	0.346	0.371	0.346	0.371

The top part of the Table reports the results of the second step, consisting in regressing the city fixed effects on log city size. In the first step, reported below, we regressed log earnings on city fixed effects and a series of controls. City size is the number of people living within 10km of the average person. The first two columns report the short-term effects. The third and fourth columns report the medium-term effects, obtained after regressing the city fixed effects augmented by the effect of average experience in the city, on city size. Controls in the first step include education, occupational categories, industry and time fixed effects, and type of contract. Clustering is implemented at the person level. Data Source: Spain's Continuous Sample of Employment Histories 2006-2013.

the dynamic benefit of moving to a large city is 1.29 percentage points in the medium-term[13]. In Column 4, the medium-term city-size earnings premium, controlling for the number of co-workers, is 0.0357. This implies that a worker moving to a city with twice the size and maintaining the same number of co-workers should expect a 2.51 percent increase after 8.9 years. Therefore, once we control for the number of co-workers, the dynamic benefits are 1.23 percentage points. This is just 0.06 percentage points lower than without controlling for the number of co-workers. This leads us to conclude that only 4.65 percent of the dynamic gains of large cities can be accounted for through the effect of large-city experience on the number of co-workers.

Notice that experience accumulated in larger cities may be positively correlated with experience accumulated in larger establishments. However, due to data limitations, we do not observe the number of co-workers throughout a worker's labor history and thus are not able to separate the effect of large-city experience from the effect of large-establishment experience. Here, we are only isolating the effect of large-city experience through an increase in the number of co-workers and find that this channel is not quantitatively important in explaining the dynamic benefits of large cities. This is consistent with the finding that the number of co-workers is increasing only slightly faster in larger cities. Finally, we find that the co-worker earnings premium remains unchanged even controlling for city-specific experience, so that a worker that experiences a doubling in the number of co-workers without moving cities or changing city-specific experience should expect a short-term 1.88 percent increase in earnings.

[13]These columns replicate the results in Table 2 of De la Roca and Puga (2017). They find very close results (a short-term 2.2 percent and a 5.1 percent after 7.7 years). The difference is likely due to the fact that we use a different time period. They use data for the years 2004 to 2009, we use 2006 to 2013.

3.5 Conclusion

This paper employs rich administrative data from Spain to document several stylized facts: the city-size earnings premium, the co-workers earnings premium, the city-size co-worker premium, and the relationship between the within-city co-worker earnings premium and city size. Once these facts are established, we ask how much of the first fact, the city-size earnings premium, can be explained by the fact that workers in large cities have more co-workers and that having more co-workers is associated with higher earnings. We find that around 30 percent of the short-term and 5 percent of the medium-term city-size earnings premium can be attributed to the fact that workers who move to larger cities also transition to better-paying larger establishments. Moreover, we find that doubling the number of co-workers and the short-term effects of doubling the size of the city are associated with similar increases in earnings of 1.88 and 1.83 percent respectively. These results highlight the importance of understanding why larger cities host larger establishments. Finally, the novel evidence we present on the negative correlation between the within-city co-worker earnings premium and city size is consistent with theories featuring imperfect labor markets and a more elastic labor supply in large cities. This could explain the part of the city-size earnings premium that is associated with differences in the establishment-size composition.

Appendix A

Appendix for "A Revealed-Preference Approach to Measuring Information Frictions in Migration Decisions"

A.1 Proofs and Additional Derivations

A.1.1 Section 1.4: Additional Derivations for the Myopic Case

Let's assume that $\lambda = 1$ and $\delta = 0$. In this case, we can write

$$v_{gmt} = b_{ndt} - \gamma_{\bar{n}ndt} + a_{ldt} - \kappa_{\bar{l}ldt} - \alpha p_{ldt} + \alpha w_{mdt}.$$

Note also that, for any two markets $m = (n, l)$ and $m' = (n', l')$ such that $n = n'$ it holds that

$$\begin{aligned} v_{gmt} - v_{gm't} = &\Big(a_{ldt} - \kappa_{\bar{l}ldt} - \alpha p_{ldt} + \alpha w_{mdt}\Big) \\ &- \Big(a_{l'dt} - \kappa_{\bar{l}l'dt} - \alpha p_{l'dt} + \alpha w_{m'dt}\Big). \end{aligned}$$

In the specific case in which the location of market m' corresponds to the prior location of workers of type g (i.e. $l' = \bar{l}$), this expression becomes:

$$\begin{aligned} v_{gmt} - v_{gm't} = &\Big(a_{ldt} - \kappa_{\bar{l}ldt} - \alpha p_{ldt} + \alpha w_{mdt}\Big) \\ &- \Big(a_{\bar{l}dt} - \kappa_{\bar{l}\bar{l}dt} - \alpha p_{\bar{l}dt} + \alpha w_{m'dt}\Big). \end{aligned}$$

Defining ξ_{gmt} as the expectational error that workers of type g make when predicting v_{gmt},

$$\xi_{gmt} \equiv v_{gmt} - \mathbb{E}[v_{gmt}|\mathcal{W}_{gt}] \tag{A.1.1}$$

we can write

$$\begin{aligned} \mathbb{E}[v_{gmt} - v_{gm't}|\mathcal{W}_{gt}] = &\Big(a_{ldt} - \kappa_{\bar{l}ldt} - \alpha p_{ldt} + \alpha w_{mdt} - \xi_{gmt}\Big) \\ &- \Big(a_{\bar{l}dt} - \kappa_{\bar{l}\bar{l}dt} - \alpha p_{\bar{l}dt} + \alpha w_{m'dt} - \xi_{gm't}\Big) \\ = &\chi_{\bar{l}ldt} + \alpha\big(w_{mdt} - w_{m'dt}\big) - \xi_{gmt} + \xi_{gm't}, \end{aligned} \tag{A.1.2}$$

where $\chi_{\bar{l}ldt}$ denotes an unobserved effect that varies by demographic type, location of the market m, prior location of workers of group g, and year. Note that, given Assumption LH5 and the definition of ξ_{gmt} in equation (A.1.1), it is the case that

$$\mathbb{E}[\xi_{gmt}|Z^*_{gt}] = 0. \tag{A.1.3}$$

Combining equations (1.14), (1.17) and (A.1.2), we can write

$$\ln\left(\frac{P(m = m_t(g)|\mathcal{W}_{it}, n = \bar{n})}{P(\bar{m} = m_t(g)|\mathcal{W}_{it}, n = \bar{n})}\right) = \chi_{\bar{l}ldt} + \alpha\big(\widehat{w}_{mdt} - \widehat{w}_{m'dt}\big) - \alpha\big(\varepsilon_{gmt} + \xi_{gmt} - \varepsilon_{gm't} - \xi_{gm't}\big). \tag{A.1.4}$$

Finally, given this estimating equation and the mean independence conditions in equations (1.16) and (A.1.3), we can derive the following conditional moment condition

$$\mathbb{E}\left[\ln\left(\frac{P(m = m_t(g)|\mathcal{W}_{it}, n = \bar{n})}{P(\bar{m} = m_t(g)|\mathcal{W}_{it}, n = \bar{n})}\right)\Big|Z^*_{gt}\right] = \mathbb{E}\big[\chi_{\bar{l}ldt} + \alpha\big(\widehat{w}_{mdt} - \widehat{w}_{m'dt}\big)\big|Z^*_{gt}\big].$$

In Section 1.6, we use this moment condition and the definition of Z^*_{gt} in equations (1.11), (1.12), and (1.13) to present estimates of the elasticity of migration with respect to wages, α.

A.1.2 Section 1.4: Additional Derivations for the Forward Looking Case

In this section, we maintain the assumption that $\lambda = 1$ but allow $\delta > 0$. Given this assumption, we can write the probability that a worker of type g chooses a market m at period t as

$$P(m = m_t(g)|\mathcal{W}_{gt}) = \frac{\exp\left(\mathbb{E}[u_{gmt}|\mathcal{W}_{gt}] + \delta\mathbb{E}[\mathsf{V}_{\mathfrak{g}t+1}|\mathcal{W}_{gt}]\right)}{\sum_{m'\in M, \mathfrak{g}'=(m',d)} \exp\left(\mathbb{E}[u_{gm't}|\mathcal{W}_{gt}] + \delta\mathbb{E}[\mathsf{V}_{\mathfrak{g}'t+1}|\mathcal{W}_{gt}]\right)},$$

where $\mathfrak{g} = (m, d)$ is the group to which a worker belong in period $t+1$ if she chooses market m in period t while being in group g, and $\mathsf{V}_{\mathfrak{g}t+1}$ is the value function at $t+1$ of a worker

in group $\mathfrak{g} = (m, d)$. Similarly, $\mathsf{V}_{\mathfrak{g}'t+1}$ is the value function at $t+1$ of a worker in group $\mathfrak{g}' = (m', d)$.

Notice that the expected ex-ante value of belonging to group $g = (\bar{m}, d)$ in year t, is given by

$$\mathbb{E}[\mathsf{V}_{gt}|\mathcal{W}_{gt}] \equiv \ln\left(\sum_{m'\in M, \mathfrak{g}'=(m',d)} \exp\left(\mathbb{E}[u_{gm't}|\mathcal{W}_{gt}] + \delta\mathbb{E}[\mathsf{V}_{\mathfrak{g}'t+1}|\mathcal{W}_{gt}]\right)\right)$$

and, therefore, we can rewrite

$$P(m = m_t(g)|\mathcal{W}_{gt}) = \frac{\exp\left(\mathbb{E}[u_{gmt}|\mathcal{W}_{gt}] + \delta\mathbb{E}[\mathsf{V}_{\mathfrak{g}t+1}|\mathcal{W}_{gt}]\right)}{\exp\left(\mathbb{E}[\mathsf{V}_{gt}|\mathcal{W}_{gt}]\right)}$$

Assuming that for any groups with same demographic type $g = (\bar{m}, d)$, $\mathfrak{g} = (m, d)$, we have $\mathcal{W}_{gt} \subseteq \mathcal{W}_{\mathfrak{g}t+1}$ (i.e. workers always gain information over time), we can use the Law of Iterated Expectations to write

$$P(m = m_t(g)|\mathcal{W}_{gt}) = \frac{\exp\left(\mathbb{E}[u_{gmt}|\mathcal{W}_{gt}] + \delta\mathbb{E}\left[\mathbb{E}[\mathsf{V}_{\mathfrak{g}t+1}|\mathcal{W}_{\mathfrak{g}t+1}]\middle|\mathcal{W}_{gt}\right]\right)}{\exp\left(\mathbb{E}[\mathsf{V}_{gt}|\mathcal{W}_{gt}]\right)}.$$

Comparing this probability to that for an alternative market m', we can write

$$\begin{aligned}\ln\left(\frac{P(m = m_t(g)|\mathcal{W}_{gt})}{P(m' = m_t(g)|\mathcal{W}_{gt})}\right) &= \mathbb{E}[u_{gmt} - u_{gm't}|\mathcal{W}_{gt}] + \delta\mathbb{E}\big[\mathbb{E}[\mathsf{V}_{\mathfrak{g}t+1}|\mathcal{W}_{\mathfrak{g}t+1}] \\ &\quad -\mathbb{E}[\mathsf{V}_{\mathfrak{g}'t+1}|\mathcal{W}_{\mathfrak{g}'t+1}]\big|\mathcal{W}_{gt}\big],\end{aligned} \tag{A.1.5}$$

where $\mathfrak{g} = (m, d)$ and $\mathfrak{g}' = (m', d)$. Similarly, we can compute the probability that both a worker of type $\mathfrak{g}$ and a worker of type $\mathfrak{g}'$ select a common market m'' at period $t+1$; i.e.

$$\begin{aligned}\ln\left(P(m'' = m_{t+1}(\mathfrak{g})|\mathcal{W}_{\mathfrak{g}t+1})\right) &= \mathbb{E}[u_{\mathfrak{g}m''t+1}|\mathcal{W}_{\mathfrak{g}t+1}] + \delta\mathbb{E}\big[\mathbb{E}[\mathsf{V}_{\mathfrak{g}''t+2}|\mathcal{W}_{\mathfrak{g}''t+2}]\big|\mathcal{W}_{\mathfrak{g}t+1}\big] \\ &\quad -\mathbb{E}[\mathsf{V}_{\mathfrak{g}t+1}|\mathcal{W}_{\mathfrak{g}t+1}],\end{aligned}$$

and

$$\begin{aligned}\ln\big(P(m''=m_{t+1}(\mathfrak{g}')|\mathcal{W}_{\mathfrak{g}'t+1})\big) &= \mathbb{E}[u_{\mathfrak{g}'m''t+1}|\mathcal{W}_{\mathfrak{g}'t+1}]+\delta\mathbb{E}\big[\mathbb{E}[\mathsf{V}_{\mathfrak{g}''t+2}|\mathcal{W}_{\mathfrak{g}''t+2}]\big|\mathcal{W}_{\mathfrak{g}'t+1}\big]\\ &\quad -\mathbb{E}[\mathsf{V}_{\mathfrak{g}'t+1}|\mathcal{W}_{\mathfrak{g}'t+1}],\end{aligned}$$

where $\mathfrak{g}''=(m'',d)$ corresponds to the group at period $t+2$ of a worker that was of type g at period t, that chose market m at period t and market m'' at period $t+1$. It is also the group at period $t+2$ of a worker that was of type g at period t and that chose market m' at period t and market m'' at period $t+1$.

By rearranging terms in the previous two equations, we can write

$$\begin{aligned}\mathbb{E}[\mathsf{V}_{\mathfrak{g}t+1}|\mathcal{W}_{\mathfrak{g}t+1}] &= \mathbb{E}[u_{\mathfrak{g}m''t+1}|\mathcal{W}_{\mathfrak{g}t+1}]+\delta\mathbb{E}\big[\mathbb{E}[\mathsf{V}_{\mathfrak{g}''t+2}|\mathcal{W}_{\mathfrak{g}''t+2}]\big|\mathcal{W}_{\mathfrak{g}t+1}\big]\\ &\quad -\ln\big(P(m''=m_{t+1}(\mathfrak{g})|\mathcal{W}_{\mathfrak{g}t+1})\big),\end{aligned}$$

and

$$\begin{aligned}\mathbb{E}[\mathsf{V}_{\mathfrak{g}'t+1}|\mathcal{W}_{\mathfrak{g}'t+1}] &= \mathbb{E}[u_{\mathfrak{g}'m''t+1}|\mathcal{W}_{\mathfrak{g}'t+1}]+\delta\mathbb{E}\big[\mathbb{E}[\mathsf{V}_{\mathfrak{g}''t+2}|\mathcal{W}_{\mathfrak{g}''t+2}]\big|\mathcal{W}_{\mathfrak{g}'t+1}\big]\big]\\ &\quad -\ln\big(P(m''=m_{t+1}(\mathfrak{g}')|\mathcal{W}_{\mathfrak{g}'t+1})\big),\end{aligned}$$

We can combine the last two expressions to obtain

$$\begin{aligned}&\mathbb{E}[\mathsf{V}_{\mathfrak{g}t+1}|\mathcal{W}_{\mathfrak{g}t+1}]-\mathbb{E}[\mathsf{V}_{\mathfrak{g}'t+1}|\mathcal{W}_{\mathfrak{g}'t+1}]\\ =\;&\mathbb{E}[u_{\mathfrak{g}m''t+1}|\mathcal{W}_{\mathfrak{g}t+1}]-\mathbb{E}[u_{\mathfrak{g}'m''t+1}|\mathcal{W}_{\mathfrak{g}'t+1}]\\ &+\delta\mathbb{E}\big[\mathbb{E}[\mathsf{V}_{\mathfrak{g}''t+2}|\mathcal{W}_{\mathfrak{g}''t+2}]\big|\mathcal{W}_{\mathfrak{g}t+1}\big]-\delta\mathbb{E}\big[\mathbb{E}[\mathsf{V}_{\mathfrak{g}''t+2}|\mathcal{W}_{\mathfrak{g}''t+2}]\big|\mathcal{W}_{\mathfrak{g}'t+1}\big]\big]\\ &-\big(\ln\big(P(m''=m_{t+1}(\mathfrak{g})|\mathcal{W}_{\mathfrak{g}t+1})\big)-\ln\big(P(m''=m_{t+1}(\mathfrak{g}')|\mathcal{W}_{\mathfrak{g}'t+1})\big)\big)\end{aligned}$$

Plugging this expression into that in equation (A.1.6), we obtain

$$\begin{aligned}&\ln\big(P(m=m_t(g)|\mathcal{W}_{gt})\big)-\ln\big(P(m'=m_t(g)|\mathcal{W}_{gt})\big)\\ =\;&\mathbb{E}[u_{gmt}-u_{gm't}|\mathcal{W}_{gt}]\end{aligned}$$

$$
\begin{aligned}
&+\delta\mathbb{E}\big[\mathbb{E}[u_{\mathfrak{g}m''t+1}|\mathcal{W}_{\mathfrak{g}t+1}] - \mathbb{E}[u_{\mathfrak{g}'m''t+1}|\mathcal{W}_{\mathfrak{g}'t+1}]\big|\mathcal{W}_{gt}\big] \\
&+\delta\mathbb{E}\left[\delta\big(\mathbb{E}\big[\mathbb{E}[\mathsf{V}_{\mathfrak{g}''t+2}|\mathcal{W}_{\mathfrak{g}''t+2}]\big|\mathcal{W}_{\mathfrak{g}t+1}\big] - \mathbb{E}\big[\mathbb{E}[\mathsf{V}_{\mathfrak{g}''t+2}|\mathcal{W}_{\mathfrak{g}''t+2}]\big|\mathcal{W}_{\mathfrak{g}'t+1}\big]\big)\big|\mathcal{W}_{gt}\right] \\
&-\delta\mathbb{E}\left[\ln\big(P(m''=m_{t+1}(\mathfrak{g})|\mathcal{W}_{\mathfrak{g}t+1}\big) - \ln\big(P(m''=m_{t+1}(\mathfrak{g}')|\mathcal{W}_{\mathfrak{g}'t+1}\big)\big|\mathcal{W}_{gt}\right]
\end{aligned}
$$

Taking into account that $\mathcal{W}_{gt} \subseteq \mathcal{W}_{\mathfrak{g}'t+1} \subseteq \mathcal{W}_{\mathfrak{g}''t+2}$ for any choices m' and m'' and $\mathfrak{g}' = (m', d)$, $\mathfrak{g}'' = (m'', d)$, we can apply the Law of Iterated Expectations to simplify the previous expression:

$$
\begin{aligned}
& \ln\big(P(m=m_t(g)|\mathcal{W}_{gt})\big) - \ln\big(P(m'=m_t(g)|\mathcal{W}_{gt})\big) \\
= \;& \mathbb{E}[u_{gmt} - u_{gm't}|\mathcal{W}_{gt}] \\
& +\delta\mathbb{E}\left[u_{\mathfrak{g}m''t+1} - u_{\mathfrak{g}'m''t+1}\big|\mathcal{W}_{gt}\right] \\
& +\delta^2\mathbb{E}\left[\mathsf{V}_{\mathfrak{g}''t+2} - \mathsf{V}_{\mathfrak{g}''t+2}\big|\mathcal{W}_{gt}\right] \\
& -\delta\mathbb{E}\Bigg[\ln\big(P(m''=m_{t+1}(\mathfrak{g})|\mathcal{W}_{\mathfrak{g}t+1}\big) \\
& -\ln\big(P(m''=m_{t+1}(\mathfrak{g}')|\mathcal{W}_{\mathfrak{g}'t+1}\big)\big|\mathcal{W}_{gt}\Bigg]
\end{aligned}
\tag{A.1.6}
$$

Given the definition of worker groups in any period $t+2$ as depending exclusively on the demographic type of the worker (which is exogenous) and the market chosen by the worker at period $t+1$, it holds that the differences in values at $t+2$ is equal to zero, simplifying the third term of the right side of the above expression. For every market m, group $g = (\bar{m}, d)$, $\mathfrak{g} = (m, d)$ and period t, we define an expectational error term,

$$
\begin{aligned}
\xi_{gmt} \;\equiv\; & u_{gmt} + \delta u_{\mathfrak{g}m''t+1} - \delta\ln\big(P(m''=m_{t+1}(\mathfrak{g})|\mathcal{W}_{\mathfrak{g}t+1}\big) \\
& -\mathbb{E}\left[u_{gmt} + \delta u_{\mathfrak{g}m''t+1} - \delta\ln\big(P(m''=m_{t+1}(\mathfrak{g})|\mathcal{W}_{\mathfrak{g}t+1}\big)\big|\mathcal{W}_{gt}\right].
\end{aligned}
$$

Given this definition of expectational error, we can simplify equation (A.1.7) as

$$
\begin{aligned}
\ln\left(\frac{P(m=m_t(g)|\mathcal{W}_{gt})}{P(m'=m_t(g)|\mathcal{W}_{gt})}\right) + \delta\ln\left(\frac{P(m''=m_{t+1}(\mathfrak{g})|\mathcal{W}_{\mathfrak{g}t+1})}{P(m''=m_{t+1}(\mathfrak{g}')|\mathcal{W}_{\mathfrak{g}'t+1})}\right) \;=\; & u_{gmt} - u_{gm't} \\
& +\delta(u_{\mathfrak{g}m''t+1} - u_{\mathfrak{g}'m''t+1})
\end{aligned}
$$

$$-\xi_{gmt} + \xi_{gm't}. \quad \text{(A.1.7)}$$

Although equation (A.1.7) holds for any markets m, m' and m'', we focus for each group $g = (\bar{m}, d)$ with $\bar{m} = (\bar{n}, \bar{l})$, on the following markets

$$m = (\bar{n}, l),$$
$$m' = (\bar{n}, \bar{l}),$$
$$m'' = (\bar{n}, l),$$

so that $m'' = m$ and $m' = \bar{m}$, and we effectively compare the discounted probabilities of taking the path $\bar{l} \to l \to l$ versus $\bar{l} \to \bar{l} \to l$. Namely, we compare individuals who start in the same location $\bar{l}$ at end of $t-1$ (and same sector and demographic type), and end up in location l in period $t+1$, either after moving to location l in period t or $t+1$.

For this choice of markets,

$$\begin{aligned} u_{gmt} - u_{gm't} &= \big(a_{ldt} - \kappa_{\bar{l}ldt} - \alpha p_{ldt} + \alpha w_{mdt}\big) \\ &\quad - \big(a_{\bar{l}dt} - \kappa_{\bar{l}\bar{l}dt} - \alpha p_{\bar{l}dt} + \alpha w_{m'dt}\big), \end{aligned} \quad \text{(A.1.8)}$$

and

$$u_{\mathfrak{g}m''t+1} - u_{\mathfrak{g}'m''t+1} = -\kappa_{\bar{l}ldt+1} + \kappa_{\bar{l}\bar{l}dt+1} \quad \text{(A.1.9)}$$

Therefore, combining equations (A.1.7), (A.1.8), and (A.1.9), we can rewrite the former as

$$\begin{aligned} \ln\left(\frac{P(m = m_t(g)|\mathcal{W}_{gt})}{P(m' = m_t(g)|\mathcal{W}_{gt})}\right) + \delta \ln\left(\frac{P(m'' = m_{t+1}(\mathfrak{g})|\mathcal{W}_{\mathfrak{g}t+1})}{P(m'' = m_{t+1}(\mathfrak{g}')|\mathcal{W}_{\mathfrak{g}'t+1})}\right) &= \chi_{\bar{l}ldt} + \alpha\left(w_{mdt} - w_{m'dt}\right) \\ &\quad - \xi_{gmt} + \xi_{gm't}. \end{aligned} \quad \text{(A.1.10)}$$

where $\chi_{\bar{l}ldt}$ denotes an unobserved effect that varies by demographic type, location of the market m, prior location of workers of group g, and year. Note that, given Assumption LH5

and the definition of ξ_{gmt} in equation (A.1.1), it is the case that

$$\mathbb{E}[\xi_{gmt}|Z^*_{gt}] = 0. \tag{A.1.11}$$

Combining equations (1.14) and (A.1.10), we can further write

$$\begin{aligned}\ln\left(\frac{P(m=m_t(g)|\mathcal{W}_{gt})}{P(m'=m_t(g)|\mathcal{W}_{gt})}\right)+\delta\ln\left(\frac{P(m''=m_{t+1}(\mathfrak{g})|\mathcal{W}_{\mathfrak{g}t+1})}{P(m''=m_{t+1}(\mathfrak{g}')|\mathcal{W}_{\mathfrak{g}'t+1})}\right) &= \chi_{\bar{l}ldt}+\alpha\left(\widehat{w}_{mdt}-\widehat{w}_{m'dt}\right)\\ &\quad -\alpha\big(\varepsilon_{mdt}-\varepsilon_{m'dt}\big)\\ &\quad -\xi_{gmt}+\xi_{gm't}. \end{aligned}\tag{A.1.12}$$

Finally, given this estimating equation and the mean independence conditions in equations (1.16) and (A.1.11), we can derive the following moment condition

$$\begin{aligned}&\mathbb{E}\left[\ln\left(\frac{P(m=m_t(g)|\mathcal{W}_{gt})}{P(m'=m_t(g)|\mathcal{W}_{gt})}\right)+\delta\ln\left(\frac{P(m''=m_{t+1}(\mathfrak{g})|\mathcal{W}_{\mathfrak{g}t+1})}{P(m''=m_{t+1}(\mathfrak{g}')|\mathcal{W}_{\mathfrak{g}'t+1})}\right)\Bigg|Z^*_{gt}\right]\\ =\;&\mathbb{E}\big[\chi_{\bar{l}ldt}+\alpha\left(\widehat{w}_{mdt}-\widehat{w}_{m'dt}\right)\big|Z^*_{gt}\big].\end{aligned}$$

In Section 1.6, we use this moment condition and the definition of Z^*_{gt} in equations (1.11), (1.12), and (1.13) to present estimates of the elasticity of migration with respect to wages, α.

A.1.3 Individual-Specific Information Sets - Partial Identification with Moment Inequalities

In this section, we present a framework that allows for the partial identification of the migration elasticity under more general assumptions than the ones maintained in the main text. Particularly, we allow information sets be specific to each *individual* and time period. We derive the revealed- preference and odds- based moment inequalities that provide identification. To simplify the exposition, we abstract from the heterogeneity across sectors. Allowing agents to also choose sectors would be a straightforward extension.

Proposition 4. *For any pair of regions j, k, for any other pair of regions ℓ, ℓ', for any time period t, for any individuals i and i', for any random vector $Z_{ii'\ell t}$ whose distribution conditional on $\mathcal{I}_{i\ell t} \cap \mathcal{I}_{i'\ell t}$ is degenerate, i.e. for $Z_{i\ell t}$ which belongs to both i' and i's information sets, the following inequalities can be used for identification of the parameter ν:*

- *Revealed-Preference Inequality*

$$\begin{aligned} & \nu 2 \ln 2 \mathbb{E}\left[\Psi(Z_{ii'\ell t}) d_{i\ell t} d_{i'\ell t} d_{i\ell' t+2} d_{i'\ell' t+2} \left(d_{i'\ell jt}(1-d_{i\ell kt}) + d_{i\ell kt}(1-d_{i'\ell jt})\right)\right] \\ & + \mathbb{E}\left[\Psi(Z_{ii'\ell t}) d_{i\ell t} d_{i'\ell t} d_{i\ell' t+2} d_{i'\ell' t+2} d_{i\ell kt} d_{i'\ell jt} \Gamma_{ii'jkt}\right] \\ \geq \; & 0 \end{aligned}$$

- *Odds-Based Inequalities*

$$\begin{aligned} & \nu \mathbb{E}\left[\Psi(Z_{ii'\ell t}) d_{i\ell t} d_{i'\ell t} d_{i\ell' t+2} d_{i'\ell' t+2} \left(d_{i\ell kt} + d_{i'\ell jt} - 4 d_{i\ell kt} d_{i'\ell jt}\right)\right] \\ & + \mathbb{E}\left[\Psi(Z_{ii'\ell t}) d_{i\ell t} d_{i'\ell t} d_{i\ell' t+2} d_{i'\ell' t+2} d_{i\ell kt} d_{i'\ell jt} \Gamma_{ii'jkt}\right] \\ \geq \; & 0 \end{aligned}$$

where

$$\Gamma_{ii'jkt} = \beta(\mu_{g(i)kt+1} - \mu_{g(i)jt+1} + \mu_{g(i')jt+1} - \mu_{g(i')kt+1}).$$

Dynamic Model of Migration with Imperfect Information

Model. Following Kennan and Walker (2011), consider an economy with J locations, and individual i'ĂŹs wage in location j is w_{ijt}. Migration decisions are made so as to maximize the expected discounted value of lifetime income, net of moving costs. Each individual belongs to an education, gender, age, which we treat as constant over time. Denote by $g(i)$ the group of i.

The wage of individual i in location j in year t is specified as

$$w_{ijt} = \mu_{g(i)jt} + \eta_i + \varepsilon_{ijt}, \tag{A.1.13}$$

where $\mu_{g(i)jt}$ is the mean wage of individuals of i's group, in location j at t, η_i is an individual effect that is fixed across locations, and ε_{ijt} is a transient effect. We assume that η, ν and ε are iid across individuals and locations.

Locations are endowed with a vector of specific amenities a_{jt} which are valued by individuals, and may be unobserved by the researcher. We also include in this vector of amenities the price level in location j at period t.

The utility flow in period t for an individual i who chooses location j depends on an idiosyncratic preference draw ζ_{ijt}, which is assumed to be iid across locations and across periods, and drawn from a type-1 extreme value distribution. Let $x_{jt} = (\mu_{jt}, a_{jt})$ denote the location-specific state, $\zeta_{it} = (\zeta_{i1t}, \ldots, \zeta_{iJt})$ the vector of preference shocks, and $\omega_{ijt} = (\eta_i, \varepsilon_{ijt}, x_{jt})$ so that the state vector is $(\omega_{ijt}, \zeta_{it})$.

The decision problem can be written in recursive form as

$$v_{ijt}(\omega_{ijt}, \zeta_{it}) = w_{ijt} + a_{jt} + \max_k \left\{ \beta \mathbb{E}_{\omega'}[V_{ikt+1}(\omega')|\mathcal{I}_{ijt}] - \kappa_{jkt} + \nu \zeta_{ikt} \right\}, \tag{A.1.14}$$

where $V_{ikt+1}(\omega_{ikt+1}) \equiv \mathbb{E}_\zeta[v_{ikt+1}(\omega_{ikt+1}, \zeta_{it})]$ and $\mathcal{I}_{ijt}$ is the information set of individual i when it is in j. In particular, differences in information sets across locations can lead to different conditional distributions of z depending on the location. The bilateral term κ_{jk} represents the cost of moving from j to k. The parameter ν governs the importance of (non-wage) idiosyncratic preferences for location decisions, so that $1/\nu$ is the elasticity of migration flows with respect to wages. In this case, following McFadden (1974) and Rust (1987), we have

$$V_{ijt}(\omega_{ijt}) = \bar{\gamma} + w_{ijt} + a_{jt} + \nu \log \sum_{k=1}^{J} \exp\left(\frac{1}{\nu} \left(\beta \mathbb{E}_{\omega'}[V_{ikt+1}(\omega')|\mathcal{I}_{ijt}] - \kappa_{jkt} \right) \right), \tag{A.1.15}$$

where $\bar{\gamma}$ is the Euler constant. Let $\rho_{ijkt}(\eta_i, \varepsilon_{ijt})$ be the probability that individual i chooses k (before ζ_{it} is realized) when the current location is j and the state is $(\eta_i, \varepsilon_{ijt})$. Note that since ε_{ijt} is iid across time and locations, its current value does not affect future payoffs and hence does not affect migration probabilities. Similarly, the individual fixed effect η_i does not affect migration probabilities.

The migration probabilities for individual i to any location k from j can be written as a function of the location state x so that $\omega = (\eta, \varepsilon, x)$,

$$\rho_{ijkt}(x) = \frac{\exp\left(\frac{1}{\nu}\left(\beta \mathbb{E}_{\omega'}[V_{ikt+1}|\mathcal{I}_{ijt}] - \kappa_{jkt}\right)\right)}{\sum_{l=1}^{J} \exp\left(\frac{1}{\nu}\left(\beta \mathbb{E}_{\omega'}[V_{ilt+1}|\mathcal{I}_{ijt}] - \kappa_{jlt}\right)\right)},$$

or, using (A.1.15) evaluated at any $\eta_i = \eta$, $\varepsilon_{ijt} = \varepsilon$,

$$\rho_{ijkt}(x) = \exp\left(\frac{1}{\nu}\left(\beta \mathbb{E}_{\omega'}[V_{ikt+1}|\mathcal{I}_{ijt}] - \kappa_{jkt} - V_{ijt}(\omega) + \bar{\gamma} + w_{ijt} + a_{jt}\right)\right). \tag{A.1.16}$$

Comparing Migration Paths. For each individual i in location ℓ at period t, we will write two different expected discounted sum of payoffs, each corresponding to different migration paths:

1. Move to k at t, then follow the path of location choices which is optimal after having moved to k at t (gives expected value $V_{i\ell kt}$).

2. Move to j at t, then follow the path of location choices which is optimal after having stayed in k at t (gives expected $W_{i\ell jt,k}$).

Denote by $d(\ell kt)_{ilt'}$ the dummy variable equal to 1 if it is optimal for individual i to choose location l at time $t' > t$, if at time t a move from ℓ to k was chosen. If at time $t'-1$ the optimal location was $l(t'-1)$, then the optimal location at time t' is chosen given the information $\mathcal{I}_{il(t'-1)t'}$.

The expected discounted sum of payoffs $V_{i\ell kt}$ for individual i in location ℓ at t conditional on choosing to move to k at t (so that the observed action is $d_{i\ell kt} = 1$), and choosing the optimal path in every period t' thereafter is

$$\begin{aligned}
& V_{i\ell kt} \\
= \; & w_{i\ell t} + a_{\ell t} - \kappa_{\ell kt} + \nu\zeta_{ikt} + \beta \mathbb{E}[w_{ikt+1} + a_{kt+1}|\mathcal{I}_{i\ell t}] \\
& + \beta \sum_{l=1}^{J} \mathbb{E}[d(\ell kt)_{ilt+1} \left(\beta\left(w_{ilt+2} + a_{lt+2}\right) - \kappa_{klt+1} + \nu\zeta_{ilt+1}\right)|\mathcal{I}_{i\ell t}] \\
& + \sum_{t'=t+2} \beta^{t'-t} \sum_{l=1}^{J} \mathbb{E}\left[d(\ell kt)_{ilt'}\left(\beta\left(w_{ilt'+1} + a_{lt'+1}\right) + \nu\zeta_{ilt'} - \sum_{l'}^{J} d(\ell kt)_{il't'-1}\kappa_{l'lt'}\right)|\mathcal{I}_{i\ell t}\right].
\end{aligned}$$

Second, the expected discounted sum of payoffs of individual i in ℓ at t conditional on choosing to move to j at t, but choosing in every period $t' > t$ the path that would have been optimal if i had moved to k at t,

$$
\begin{aligned}
& W_{i\ell jt,k} \\
= \; & w_{i\ell t} + a_{\ell t} - \kappa_{\ell jt}\nu\zeta_{ijt} + \beta\mathbb{E}[w_{ijt+1} + a_{jt+1}|\mathcal{I}_{i\ell t}] \\
& + \beta \sum_{l=1}^{J} \mathbb{E}[d(\ell jt)_{ilt+1} \left(\beta \left(w_{ilt+2} + a_{lt+2}\right) - \kappa_{jlt+1} + \nu\zeta_{ilt+1}\right)|\mathcal{I}_{i\ell t}] \\
& + \sum_{t'=t+2} \beta^{t'-t} \sum_{l=1}^{J} \mathbb{E}\left[d(\ell jt)_{ilt'} \left(\beta \left(w_{ilt'+1} + a_{lt'+1}\right) + \nu\zeta_{ilt'} - \sum_{l'}^{J} d(\ell jt)_{il't'-1}\kappa_{l'lt'} \right) |\mathcal{I}_{i\ell t} \right].
\end{aligned}
$$

Since $V_{i\ell kt}$ is the expected value of choosing the optimal path after having moved to k at t, by definition it must offer a larger expected value than moving to k then following the optimal path associated to having moved to j at t,

$$V_{i\ell kt} - W_{i\ell kt,j} \geq 0, \qquad \forall \ell, j, k, t. \tag{A.1.17}$$

Revealed-Preference Moment Inequalities

First Moment Inequality for i. Fix a pair of locations j and k, and select an individual i located in any location ℓ at t, such that k belongs to i's consideration set. Denote by ℓ' the location of i at $t+2$, which is unrestricted.

This individual i will be compared to another individual i' who needs to satisfy the condition of being located in the same region as i at t, namely ℓ, will move to j at $t+1$, and then move to the same location at i at $t+2$, namely ℓ'. In the model, the infinite support of the distribution of idiosyncratic taste shocks ensures that for any i, there will always exist an individual i'.

In the data however, if no individual i' satisfies the condition, then the conditioning will impose some selection on the individuals i to whom the inequality applies. In order to avoid this selection, we restrict the set of regions to a set of "well connected" regions, forming a

dense network of migration between every location at every period. This restriction will mitigate, and arguably even avoid this selection issue.

Define the individual location dummy $d_{i\ell t}$ indicating that individual i is located in ℓ at t as

$$d_{i\ell t} \equiv \mathbb{1}\{i \text{ is in } \ell \text{ at } t\}.$$

Define the individual migration dummy $d_{i\ell kt}$ indicating that individual i moved from ℓ to k at t as

$$d_{i\ell kt} \equiv \mathbb{1}\{V_{i\ell kt} - V_{i\ell lt} \geq 0, \forall l \in \mathcal{C}_{i\ell t}\},$$

where $\mathcal{C}_{i\ell t}$ is the consideration set of i at t. This implies that for every individual i in ℓ at t, for any location k in the consideration set of i, we have

$$d_{i\ell kt}\left(V_{i\ell kt} - V_{i\ell jt}\right) \geq 0. \tag{A.1.18}$$

This exploits the fact that $\mathbb{1}\{V_{i\ell kt} - V_{i\ell jt} \geq 0\}$ is a necessary condition for $d_{i\ell kt} = 1$.

Note that rewriting (A.1.17) with origin ℓ, second destination j, and reference k for the continuation path, we have

$$V_{i\ell jt} - W_{i\ell jt,k} \geq 0. \tag{A.1.19}$$

Combining equations (A.1.18) with (A.1.19), we obtain

$$d_{i\ell kt}\left(V_{i\ell kt} - W_{i\ell jt,k}\right) \geq 0. \tag{A.1.20}$$

Note that $V_{i\ell kt} - W_{i\ell jt,k}$ can be decomposed as a sum of a "payoff" and "selection" term,

$$V_{i\ell kt} - W_{i\ell jt,k} = \Delta_{i\ell kjt} + \nu(\zeta_{ikt} - \zeta_{ijt}), \tag{A.1.21}$$

where

$$\begin{aligned}\Delta_{i\ell kjt} &= \mathbb{E}[-\kappa_{\ell kt} + \kappa_{\ell jt}|\mathcal{I}_{i\ell t}] \\ &\quad +\beta\mathbb{E}[w_{ikt+1} + a_{kt+1} - w_{ijt+1} - a_{jt+1}|\mathcal{I}_{i\ell t}]\end{aligned}$$

$$
\begin{aligned}
&+\beta \sum_l \mathbb{E}[d(\ell kt)_{ilt+1}(\kappa_{jlt+1} - \kappa_{klt+1})|\mathcal{I}_{i\ell t}] \\
= \quad & -\kappa_{\ell kt} + \kappa_{\ell jt} \\
&+\beta \left(\mu_{g(i)kt+1} + a_{kt+1} + \eta_i + \varepsilon_{ikt+1} - \mu_{g(i)jt+1} - a_{jt+1} - \eta_i - \varepsilon_{ijt+1}\right) \\
&+\beta \sum_l d(\ell kt)_{ilt+1}(\kappa_{jlt+1} - \kappa_{klt+1}) \\
&-\beta \left(\chi^w_{i\ell kt} + \chi^a_{i\ell kt} - \chi^w_{i\ell jt} - \chi^a_{i\ell jt} + \chi^d_{i\ell kjt}\right),
\end{aligned}
$$

where we define the expectational errors on wages, amenities and the term containing migration costs as

$$
\chi^w_{i\ell kt} = w_{ikt+1} - \mathbb{E}[w_{ikt+1}|\mathcal{I}_{i\ell t}]. \tag{A.1.22}
$$

$$
\begin{aligned}
\chi^d_{i\ell kjt} \quad = \quad & \left(-\kappa_{\ell kt} + \kappa_{\ell jt}\right)/\beta + \sum_l d(\ell kt)_{ilt+1}(\kappa_{jlt+1} - \kappa_{klt+1}) \\
& -\mathbb{E}\left[\left(-\kappa_{\ell kt} + \kappa_{\ell jt}\right)/\beta + \sum_l d(\ell kt)_{ilt+1}(\kappa_{jlt+1} - \kappa_{klt+1})|\mathcal{I}_{i\ell t}\right].
\end{aligned}
$$

Although $V_{i\ell kt} - W_{i\ell jt,k}$ is the difference between values along entire paths of migration, it only depends on $t+1$ payoffs. Indeed, both $V_{i\ell kt}$ and $W_{i\ell jt,k}$ correspond to following from $t+1$ onwards the migration path which is optimal after moving from ℓ to k at t. The only difference between these two expected discounted sum of payoffs relates to what happens in periods t and on the transition to $t+1$. The first two terms reflect that in one case the migration cost to k is incurred in the first period and the next period wage is in k, while in the other case the migration cost is to j and the next wage is in j. The last term shows that before the two paths join, migration to the optimal location l at $t+1$ must occur, and will entail in one case a move from k to l, while the move will be from j to l in the other case.

Rewriting (A.1.20), we obtain for any i in ℓ at t

$$
d_{i\ell kt}\left(\Delta_{i\ell kjt} + \nu(\zeta_{ikt} - \zeta_{ijt})\right) \geq 0. \tag{A.1.23}
$$

We argued above that for this individual i, there exists another individual i' such that i' is also in ℓ at t, and i' moves to j at $t+1$ and moves to the same place ℓ' as i in $t+2$. We

can therefore write inequality (A.1.23) conditional on the existence of such an individual i'. We can summarize the conditioning set that we impose on i as

$$\Omega_{ii'\ell\ell'jt} = \{d_{i\ell t} = 1, \mathcal{I}_{i\ell t}, d_{i'\ell t} = 1, \mathcal{I}_{i'\ell t}, d_{i'\ell jt} = 1, d_{i\ell' t+2} = 1, d_{i'\ell' t+2} = 1\}, \tag{A.1.24}$$

And since the inequality above holds for any i conditional on $\Omega_{ii'\ell\ell'jt}$, it must also hold in expectation,

$$\mathbb{E}\left[d_{i\ell kt}\left(\Delta_{i\ell kjt} + \nu(\zeta_{ikt} - \zeta_{ijt})\right)|\Omega_{ii'\ell\ell'jt}\right] \geq 0, \tag{A.1.25}$$

where the expectation is taken over all possible realizations of random variables entering i's decision rule, including preference shocks, conditinal on i being in ℓ at t with information set $\mathcal{I}_{i\ell t}$, and on i' being in ℓ at t with information set $\mathcal{I}_{i'\ell t}$, moving to j at t and moving to the same location as i in $t+2$.

We now focus on the second part of the expectation, which reflects selection on unobserved preferences, and look for an upper bound on this term. The observed decision to move to k is the result of a comparison over many alternative location choices. In order to get a bound, we bring the comparison back to a binary decision between j and k.

To obtain this, first note that conditional on i's information set and i being in ℓ at t, the preference shocks have mean zero,

$$\mathbb{E}\left[\zeta_{ikt} - \zeta_{ijt}|d_{i\ell t} = 1, \mathcal{I}_{i\ell t}\right] = 0. \tag{A.1.26}$$

This is where the initial choice of regions j and k ensures that conditioning on the existence of i' satisfying the conditions spelled out in $\Omega_{ii'\ell\ell'jt}$ does not bring about any selection in the set of individuals i. As a result, adding the conditioning on i''s existence, location, mobility choice and information set does not affect the distribution of preference draws of individual i. Therefore,

$$\mathbb{E}\left[\zeta_{ikt} - \zeta_{ijt}|\Omega_{ii'\ell\ell'jt}\right] = 0. \tag{A.1.27}$$

Therefore, we can "switch" the conditioning on *not* moving to k in order to find the upper bound, and apply the law of iterated expectation to write

$$\begin{aligned}
& \mathbb{E}\left[d_{i\ell kt}\left(\zeta_{ikt}-\zeta_{ijt}\right)|\Omega_{ii'\ell\ell'jt}\right] \\
= \; & -\mathbb{E}\left[(1-d_{i\ell kt})(\zeta_{ikt}-\zeta_{ijt})|\Omega_{ii'\ell\ell'jt}\right] \\
= \; & \mathbb{E}\left[(1-d_{i\ell kt})\mathbb{E}[\zeta_{ijt}-\zeta_{ikt}|d_{i\ell t}=1,d_{i\ell kt}=0,\mathcal{I}_{i\ell t}]|\Omega_{ii'\ell\ell'jt}\right],
\end{aligned}$$

where in the last line we retained only the relevant conditioning event,

$$\begin{aligned}
& \mathbb{E}[\zeta_{ijt}-\zeta_{ikt}|d_{i\ell t}=1,d_{i\ell kt}=0,\mathcal{I}_{i\ell t},d_{i'\ell t}=1,\mathcal{I}_{i'\ell t},d_{i'\ell jt}=1,d_{i\ell' t+2}=1,d_{i'\ell' t+2}=1] \\
= \; & \mathbb{E}[\zeta_{ijt}-\zeta_{ikt}|d_{i\ell t}=1,d_{i\ell kt}=0,\mathcal{I}_{i\ell t}].
\end{aligned}$$

Now note that in the context of multiple alternatives, not choosing alternative k does not mean that alternative j is preferred to k, but in the context of binary alternatives it does. Therefore, the expected difference of the shocks is larger in the context of binary choice

$$\begin{aligned}
& \mathbb{E}[\zeta_{ijt}-\zeta_{ikt}|d_{i\ell t}=1,d_{i\ell kt}=0,\mathcal{I}_{i\ell t}] \\
\leq \; & \mathbb{E}[\zeta_{ijt}-\zeta_{ikt}|\mathcal{C}_{i\ell t}=\{j,k\},d_{i\ell t}=1,d_{i\ell kt}=0,\mathcal{I}_{i\ell t}],
\end{aligned}$$

where $\mathcal{C}_{i\ell t}$ is the consideration set of individual i when it is in location ℓ at t. We can then rewrite the bound as

$$\begin{aligned}
& \mathbb{E}[\zeta_{ijt}-\zeta_{ikt}|\mathcal{C}_{i\ell t}=\{j,k\},d_{i\ell t}=1,d_{i\ell kt}=0,\mathcal{I}_{i\ell t}] \\
= \; & \mathbb{E}[\zeta_{ijt}-\zeta_{ikt}|V_{i\ell jt}-V_{i\ell kt}\geq 0,d_{i\ell t}=1,\mathcal{I}_{i\ell t}] \\
= \; & \mathbb{E}[\zeta_{ijt}-\zeta_{ikt}|\zeta_{ijt}-\zeta_{ikt}\geq \tilde{\Delta}_{i\ell kjt}/\nu,d_{i\ell t}=1,\mathcal{I}_{i\ell t}],
\end{aligned}$$

where we defined

$$\tilde{\Delta}_{i\ell kjt}=V_{i\ell kt}-V_{i\ell jt}+\nu(\zeta_{ijt}-\zeta_{ikt}). \tag{A.1.28}$$

Since the preference draws ζ_{ijt} are iid type-1 extreme value shocks, the difference $\zeta_{ijt}-\zeta_{ikt}$ has a logistic distribution. It is easy to show that the left-truncated mean of a logistic distribution with mean 0 and variance σ^2 is

$$g(x) = \mathbb{E}[z|z \geq x] = x + (1 + e^x) \ln\left(1 + e^{-x}\right). \tag{A.1.29}$$

It is also easy to show that $g(\cdot)$ is increasing and convex. Hence we can bound the selection term by

$$\mathbb{E}\left[d_{i\ell kt}\left(\zeta_{ikt} - \zeta_{ijt}\right)|\Omega_{ii'\ell\ell'jt}\right] \leq \mathbb{E}\left[(1 - d_{i\ell kt})g\left(\tilde{\Delta}_{i\ell kjt}/\nu\right)|\Omega_{ii'\ell\ell'jt}\right]. \tag{A.1.30}$$

Plugging the upper bound into inequality (A.1.25),

$$\mathbb{E}\left[d_{i\ell kt}\Delta_{i\ell kjt} + \nu(1 - d_{i\ell kt})g\left(\tilde{\Delta}_{i\ell kjt}/\nu\right)|\Omega_{ii'\ell\ell'jt}\right] \geq 0. \tag{A.1.31}$$

The payoff difference $\tilde{\Delta}_{i\ell kjt}$ cannot be computed as a function of t and $t+1$ payoffs only as it is the difference between two whole migration paths. However, we can bound it above by $\Delta_{i\ell kjt}$ and use the fact that $g(\cdot)$ is increasing to replace it in the inequality. Using (A.1.21), (A.1.28) and (A.1.19), we get

$$\begin{aligned}
\Delta_{i\ell kjt} - \tilde{\Delta}_{i\ell kjt} &= V_{i\ell kt} - W_{i\ell jt,k} - (V_{i\ell kt} - V_{i\ell jt}) \\
&= V_{i\ell jt} - W_{i\ell jt,k} \\
&\geq 0.
\end{aligned}$$

This gives the following inequality

$$\mathbb{E}\left[d_{i\ell kt}\Delta_{i\ell kjt} + \nu(1 - d_{i\ell kt})g\left(\Delta_{i\ell kjt}/\nu\right)|\Omega_{ii'\ell\ell'jt}\right] \geq 0, \tag{A.1.32}$$

Note from the definition of $\Delta_{i\ell kjt}$ that it is a function of variables that are all part of $\mathcal{I}_{i\ell t}$. Therefore, we can divide inside the expectation by $\nu g\left(\Delta_{i\ell kjt}/\nu\right)$

$$\mathbb{E}\left[d_{i\ell kt}h\left(\Delta_{i\ell kjt}/\nu\right)+(1-d_{i\ell kt})|\Omega_{ii'\ell\ell'jt}\right]\geq 0, \tag{A.1.33}$$

where

$$h(x)=\frac{x}{g(x)}=\frac{x}{x+(1+e^{x})\ln\left(1+e^{-x}\right)}. \tag{A.1.34}$$

It can be shown easily that h is an increasing concave function. Dividing by $g\left(\Delta_{i\ell kjt}/\nu\right)$ achieves two goals. First, the definition of $\Delta_{i\ell kjt}$ involves the dummy $d(\ell kt)_{ilt+1}$, which depends on the optimal location decision following a move from ℓ to k at t, and therefore is not observed for individuals with $1-d_{i\ell kt}=1$, making the product $(1-d_{i\ell kt})g\left(\Delta_{i\ell kjt}/\nu\right)$ impossible to evaluate directly. In (A.1.33), the term $\Delta_{i\ell kjt}/\nu$ is only evaluated for agents with $d_{i\ell kt}=1$. Second, we have replaced the convex function $g(\cdot)$ by the concave function $h(\cdot)$, which has the simple property of being bounded above by any of its tangents. Since $h(0)=0$, and $h'(0)=1/(2\ln 2)$, we can bound $h(x)$ by

$$h(x)\leq\frac{1}{2\ln 2}x,\forall x\in R. \tag{A.1.35}$$

As a result, we can further simplify (A.1.33) into an expectation of linear terms

$$\mathbb{E}\left[d_{i\ell kt}\Delta_{i\ell kjt}/\nu+2\ln 2(1-d_{i\ell kjt})|\Omega_{ii'\ell\ell'jt}\right]\geq 0, \tag{A.1.36}$$

This implies that, for any random vector $Z_{ii'\ell t}$ whose distribution conditional on $\mathcal{I}_{i\ell t}\cap\mathcal{I}_{i'\ell t}$ is degenerate, i.e. for $Z_{ii'\ell t}$ which belongs to both i's and i''s information sets, and for any function $\Psi(\cdot)$ such that

$$\Psi(Z_{ii'\ell t})\geq 0,\qquad\forall Z_{ii'\ell t}, \tag{A.1.37}$$

we get the unconditional expectation

$$\mathbb{E}\left[\Psi(Z_{ii'\ell t})d_{i\ell t}d_{i'\ell t}d_{i\ell' t+2}d_{i'\ell' t+2}d_{i'\ell jt}\left(d_{i\ell kt}\Delta_{i\ell kjt}/\nu+2\ln 2(1-d_{i\ell kt})\right)\right]\geq 0. \tag{A.1.38}$$

Second Moment Inequality for i'. Note that since j and k can be any pair of locations, one can write the same equation as (A.1.38), but flipping j and k. That is, we pick individual i' in ℓ at t and ℓ' at $t+2$, and consider some other individual i in ℓ at t and ℓ' at $t+2$ and moves to k at t

$$\mathbb{E}\left[\Psi(Z_{ii'\ell t})d_{i\ell t}d_{i'\ell t}d_{i\ell' t+2}d_{i'\ell' t+2}d_{i\ell kt}\left(d_{i'\ell jt}\Delta_{i'\ell jkt}/\nu + 2\ln 2(1-d_{i'\ell jt})\right)\right] \geq 0. \quad \text{(A.1.39)}$$

Summing Inequalities. We can sum (A.1.38) and (A.1.39) to obtain

$$\begin{aligned}
&\mathbb{E}\Bigg[\Psi(Z_{ii'\ell t})d_{i\ell t}d_{i'\ell t}d_{i\ell' t+2}d_{i'\ell' t+2}\times \\
&\left(d_{i\ell kt}d_{i'\ell jt}\left(\Delta_{i\ell kjt}+\Delta_{i'\ell jkt}\right)/\nu + 2\ln 2\left(d_{i'\ell jt}(1-d_{i\ell kt})+d_{i\ell kt}(1-d_{i'\ell jt})\right)\right)\Bigg] \\
\geq\ & 0.
\end{aligned}$$

We now focus on the sum of $\Delta_{i\ell kjt}$ and $\Delta_{i'\ell jkt}$ and show how it cancels out the true values of migration costs κ_{jkt}, κ_{jjt}, and amenities a_{kt+1}, a_{jt+1}. Recall that the dummy $d(\ell kt)_{ilt+1}$ takes value 1 if and only if i who moved from ℓ to k at t finds it optimal to move to l at $t+1$. Hence, the condition that i and i' chose the same location ℓ' at $t+2$ implies that $d(\ell kt)_{ilt+1} = d(\ell jt)_{i'lt+1}$, for all l, so that

$$\begin{aligned}
\Delta_{i\ell kjt}+\Delta_{i'\ell jkt} \;=\;& -\kappa_{\ell kt}+\kappa_{\ell jt} \\
&+\beta\left(\mu_{g(i)kt+1}+a_{kt+1}+\varepsilon_{ikt+1}-\mu_{g(i)jt+1}-a_{jt+1}-\varepsilon_{ijt+1}\right) \\
&+\beta\sum_l d(\ell kt)_{ilt+1}(\kappa_{jlt+1}-\kappa_{klt+1}) \\
&-\beta\left(\chi^w_{i\ell kt}+\chi^a_{i\ell kt}-\chi^w_{i\ell jt}-\chi^a_{i\ell jt}+\chi^d_{i\ell kjt}\right) \\
&-\kappa_{\ell jt}+\kappa_{\ell kt} \\
&+\beta\left(\mu_{g(i')jt+1}+a_{jt+1}+\varepsilon_{i'kt+1}-\mu_{g(i')jt+1}-a_{kt+1}-\varepsilon_{i'jt+1}\right) \\
&+\beta\sum_l d(\ell jt)_{i'lt+1}(\kappa_{klt+1}-\kappa_{jlt+1}) \\
&-\beta\left(\chi^w_{i'\ell jt}+\chi^a_{i'\ell jt}-\chi^w_{i'\ell kt}-\chi^a_{i'\ell kt}+\chi^d_{i'\ell jkt}\right)
\end{aligned}$$

$$
\begin{aligned}
= \quad & \beta(\mu_{g(i)kt+1} - \mu_{g(i)jt+1} + \mu_{g(i')jt+1} - \mu_{g(i')kt+1}) \\
& +\beta(\varepsilon_{ikt+1} - \varepsilon_{ijt+1} + \varepsilon_{i'jt+1} - \varepsilon_{i'kt+1}) \\
& -\beta\left(\chi^w_{i\ell kt} + \chi^a_{i\ell kt} - \chi^w_{i\ell jt} - \chi^a_{i\ell jt} + \chi^d_{i\ell kjt}\right) \\
& -\beta\left(\chi^w_{i'\ell jt} + \chi^a_{i'\ell jt} - \chi^w_{i'\ell kt} - \chi^a_{i'\ell kt} + \chi^d_{i'\ell jkt}\right)
\end{aligned}
$$

Notice that in the inequality above, we form the expectation of $\Delta_{i\ell kjt}+\Delta_{i'\ell jkt}$ multiplied by a function of $Z_{ii'\ell t} \in \mathcal{I}_{i\ell t} \cap \mathcal{I}_{i'\ell t}$. Since location and migration decisions are functions of individuals' information set at the time of decision, the dummies $d_{i\ell t}$, $d_{i'\ell t}$, $d_{i\ell' t+2}$, $d_{i'\ell' t+2}$ do not provide any information on wages that was not already contained in the information set. Therefore, the dummy location and migration variables are orthogonal to the contemporaneous expectational errors.

Hence, since by definition of the expectational error, $\mathbb{E}[\chi^w_{i\ell kt}|\mathcal{I}_{i\ell t}] = 0$, it holds that for any given i' and $Z_{ii'\ell tt+2} \in \mathcal{I}_{i\ell t} \cap \mathcal{I}_{i\ell t+2}$

$$
\mathbb{E}\left[\Psi(Z_{ii'\ell t}) d_{i\ell t} d_{i'\ell t} d_{i\ell' t+2} d_{i'\ell' t+2} \chi^w_{i\ell kt}\right] = 0, \tag{A.1.40}
$$

and similarly for $\chi^a_{i\ell kt}$, $\chi^w_{i'\ell jt}$, $\chi^a_{i'\ell jt}$,$\chi^d_{i\ell kjt}$,$\chi^d_{i'\ell jkt}$ so that

$$
\mathbb{E}\left[\Psi(Z_{ii'\ell t}) d_{i\ell t} d_{i'\ell t} d_{i\ell' t+2} d_{i'\ell' t+2} \Xi_{ii'\ell jkt}\right] = 0, \tag{A.1.41}
$$

where $\Xi_{ii'\ell jkt}$ is defined as

$$
\Xi_{ii'\ell jkt} = \left(\chi^w_{i\ell kt} + \chi^a_{i\ell kt} - \chi^w_{i\ell jt} - \chi^a_{i\ell jt} + \chi^d_{i\ell kjt}\right) + \left(\chi^w_{i'\ell jt} + \chi^a_{i'\ell jt} - \chi^w_{i'\ell kt} - \chi^a_{i'\ell kt} + \chi^d_{i'\ell jkt}\right). \tag{A.1.42}
$$

Noting further that the transient shocks are by assumption orthogonal to agent+2s decisions, it holds that

$$
\mathbb{E}_i\left[\Psi(Z_{ii'\ell t}) d_{i\ell t} d_{i'\ell t} d_{i\ell' t+2} d_{i'\ell' t+2} \varepsilon_{ikt+1}\right] = 0, \tag{A.1.43}
$$

and similarly for ε_{ijt+1}, $\varepsilon_{i'kt+1}$, $\varepsilon_{i'jt+1}$. As a result, we can rewrite the inequality and get the following revealed preference inequality

$$\begin{aligned} & \mathbb{E}\Bigg[\Psi(Z_{ii'\ell t})d_{i\ell t}d_{i'\ell t}d_{i\ell' t+2}d_{i'\ell' t+2}\times \\ & \left(d_{i\ell kt}d_{i'\ell jt}\Gamma_{ii'jkt}/\nu + 2\ln 2\left(d_{i'\ell jt}(1-d_{i\ell kt}) + d_{i\ell kt}(1-d_{i'\ell jt})\right)\right)\Bigg] \\ \geq \;& 0, \end{aligned}$$

where we defined $\Gamma_{ii'jkt}$ as a function of observables only,

$$\Gamma_{ii'jkt} \;=\; \beta(\mu_{g(i)kt+1} - \mu_{g(i)jt+1} + \mu_{g(i')jt+1} - \mu_{g(i')kt+1}).$$

Since this is only a condition on the parameter ν, we can write

$$\nu \geq -\frac{\mathbb{E}\left[\Psi(Z_{ii'\ell t})d_{i\ell t}d_{i'\ell t}d_{i\ell' t+2}d_{i'\ell' t+2}d_{i\ell kt}d_{i'\ell jt}\Gamma_{ii'jkt}\right]}{2\ln 2\mathbb{E}\left[\Psi(Z_{ii'\ell t})d_{i\ell t}d_{i'\ell t}d_{i\ell' t+2}d_{i'\ell' t+2}\left(d_{i'\ell jt}(1-d_{i\ell kt}) + d_{i\ell kt}(1-d_{i'\ell jt})\right)\right]}. \quad \text{(A.1.44)}$$

Odds-Based Moment Inequalities

First Moment Inequality for i. Fix a pair of locations j and k, and select an individual i located in any location ℓ at t, such that k belongs to i's consideration set. Denote by ℓ' the location of i at $t+2$, which is unrestricted.

The definition of the random variable $d_{i\ell kt}$ implies that, for every individual i, origin location j, destination k and period t, we can write the following inequality

$$\mathbb{1}\{V_{i\ell kt} - V_{i\ell jt} \geq 0\} - d_{i\ell kt} \geq 0. \quad \text{(A.1.45)}$$

This equation exploits the fact that $\mathbb{1}\{V_{i\ell kt} - V_{i\ell jt} \geq 0\}$ is a necessary condition for $d_{i\ell kt} = 1$. Combining equations (A.1.19) and (A.1.45), we obtain

$$\mathbb{1}\{V_{i\ell kt} - W_{i\ell jt,k} \geq 0\} - d_{i\ell kt} \geq 0. \quad \text{(A.1.46)}$$

And from equation (A.1.21), we can rewrite as

$$\mathbb{1}\{\zeta_{ijt} - \zeta_{ikt} \leq \Delta_{i\ell kjt}/\nu\} - d_{i\ell kt} \geq 0. \tag{A.1.47}$$

We argued above that for this individual i, there exists another individual i' such that i' is also in ℓ at t, and i' moves to j at $t+1$ and moves to the same place ℓ' as i in $t+2$. Choosing one such i', we can summarize the conditioning set that we impose on i as in the previous section by $\Omega_{ii'\ell\ell'jt}$, and write condition (A.1.47) as an expectation

$$\mathbb{E}\left[\mathbb{1}\{\zeta_{ijt} - \zeta_{ikt} \leq \Delta_{i\ell kjt}/\nu\} - d_{i\ell kt} | \Omega_{ii'\ell\ell'jt}\right] \geq 0. \tag{A.1.48}$$

Adding the conditioning on i''s current location and information sets, as well as i''s mobility choice does not induce any selection on the set of individuals i for which this inequality is valid, and therefore does not affect the distribution of i's idiosyncratic preference shocks. Therefore the difference in preferences is distributed logistic and we can write

$$\mathbb{E}\left[\mathbb{1}\{\zeta_{ijt} - \zeta_{ikt} \leq \Delta_{i\ell kjt}/\nu\} | \Omega_{ii'\ell\ell'jt}\right] = F\left(\Delta_{i\ell kjt}/\nu\right) \tag{A.1.49}$$

And since $\Delta_{i\ell kjt}$ is a function of variables that belong to $\mathcal{I}_{i\ell t}$, we can divide the argument of the expectation by $F\left(\Delta_{i\ell kjt}/\nu\right)$ and get

$$\mathbb{E}\left[1 - d_{i\ell kt}\left(1 + \exp\left(-\Delta_{i\ell kjt}/\nu\right)\right) | \Omega_{ii'\ell\ell'jt}\right] \geq 0, \tag{A.1.50}$$

where we used

$$F(z) = \frac{1}{1+e^{-z}}, \qquad z \in \mathbb{R}, \qquad \frac{1}{F\left(z\right)} = 1 + e^{-z}. \tag{A.1.51}$$

Further, we use

$$e^z \geq 1 + z \quad \forall z \in \mathbb{R}$$

to get

$$\mathbb{E}\left[1 - d_{i\ell kt}\left(2 - \Delta_{i\ell kjt}/\nu\right) | \Omega_{ii'\ell\ell'jt}\right] \geq 0. \tag{A.1.52}$$

This implies that, for any random vector $Z_{ii'\ell t}$ whose distribution conditional on $\mathcal{I}_{i\ell t}\mathcal{I}_{i'\ell\sqcup}$ is degenerate, i.e. for $Z_{ii'\ell t}$ which belongs to both i's and i''s information sets, and for any function $\Psi(\cdot)$ such that

$$\Psi(Z_{ii'\ell t}) \geq 0, \qquad \forall Z_{ii'\ell t}, \tag{A.1.53}$$

we get the unconditional expectation

$$\mathbb{E}\left[\Psi(Z_{ii'\ell t})d_{i\ell t}d_{i'\ell t}d_{i\ell' t+2}d_{i'\ell' t+2}d_{i'\ell jt}\left(1-d_{i\ell kt}\left(2-\Delta_{i\ell kjt}/\nu\right)\right)\right] \geq 0. \tag{A.1.54}$$

Second Moment Inequality for i'. Note that since j and k can be any pair of locations, one can write the same equation as (A.1.54), but flipping j and k. That is, we pick individual i' in ℓ at t and ℓ' at $t+2$, and consider some other individual i in ℓ at t and ℓ' at $t+2$ and moves to k at t,

$$\mathbb{E}\left[\Psi(Z_{ii'\ell t})d_{i\ell t}d_{i'\ell t}d_{i\ell' t+2}d_{i'\ell' t+2}d_{i\ell kt}\left(1-d_{i'\ell jt}\left(2-\Delta_{i'\ell jkt}/\nu\right)\right)\right] \geq 0. \tag{A.1.55}$$

Summing Inequalities. We can sum (A.1.54) and (A.1.55) to obtain

$$\mathbb{E}\left[\Psi(Z_{ii'\ell t})d_{i\ell t}d_{i'\ell t}d_{i\ell' t+2}d_{i'\ell' t+2}\left(d_{i\ell kt}+d_{i'\ell jt}+d_{i\ell kt}d_{i'\ell jt}\left(\left(\Delta_{i\ell kjt}+\Delta_{i'\ell jkt}\right)/\nu\right)-4\right)\right] \geq 0. \tag{A.1.56}$$

Using the analysis of the sum of $\Delta_{i\ell kjt}$ and $\Delta_{i'\ell jkt}$ from the previous section, we can write the following odds-based inequality

$$\mathbb{E}\left[\Psi(Z_{ii'\ell t})d_{i\ell t}d_{i'\ell t}d_{i\ell' t+2}d_{i'\ell' t+2}\left(d_{i\ell kt}+d_{i'\ell jt}+d_{i\ell kt}d_{i'\ell jt}\left(\Gamma_{ii'jkt}/\nu\right)-4\right)\right] \geq 0 \tag{A.1.57}$$

where we defined $\Gamma_{ii'jkt}$ as a function of observables only,

$$\Gamma_{ii'jkt} \quad = \quad \beta(\mu_{g(i)jt+1}-\mu_{g(i)kt+1}+\mu_{g(i')kt+1}-\mu_{g(i')jt+1}).$$

A.2 Descriptive Statistics and Additional Suggestive Evidence on Information Frictions

A.2.1 Descriptive Statistics

Predicting Earnings in Each Microregion

Table A.2.1: Average Fixed Effect Estimates by Year

Year	Average Fixed Effects
1999	-1.217
2000	-1.608
2001	-1.670
2002	-0.157
2003	0.0722
2004	0.0198
2005	0.0790
2006	0.178
2007	0.212
2008	0.292
2009	0.320
2010	0.403
2011	-0.305
2012	0.160
2013	0.688
2014	0.795

Table A.2.2: Average Fixed Effect Estimates by Demographic Group

Demographic Type	Code	Average Fixed Effects
Male ; White ; Below High School	1	-0.0992
Male ; Non-White ; Below High School	2	-0.241
Female ; White ; Below High School	3	-0.400
Female ; Non-White ; Below High School	4	-0.493
Male ; White ; Above High School	5	0.405
Male ; Non-White ; Above High School	6	0.159
Female ; White ; Above High School	7	0.0545
Female ; Non-White ; Above High School	8	-0.104

Table A.2.3: Average Fixed Effect Estimates by Sector

Sector	Average Fixed Effects
Agriculture - Animal and Vegetable Products	-0.765
Silviculture and Forest exploration	-0.499
Fishery	-0.919
Coal Mining	0.573
Petroleum and Gas Extraction	1.419
Metal and radioactive Ore Mining	0.733
Nonmetallic Mineral Mining	0.017
Food Processing and Manuf.	0.008
Tobacco Product Manuf.	-0.466
Textile Products Manuf.	0.055
Apparel Accessories Manuf.	-0.436
Leather Products Manuf.	-0.570
Wood Products Manuf.	-0.375
Pulp, Paper and Paperboard Products Manuf.	0.629

Medias Publishing, Printing and Reproducing	0.340
Coal & Nuclear Combustibles Processing	0.426
Miscellaneous Chemical Product Manuf.	0.652
Plastics Product Manuf.	0.382
Nonmetallic Mineral Product Manuf.	-0.005
Nonferrous Metal Foundries	0.716
Metal Product Manuf.	0.335
Household Appliance & Equipments Manuf.	0.569
Office Machinery and Electronic Computer Manuf.	0.530
Electrical Machinery & Supplies Manuf.	0.510
Audio and Video Equipment Manuf.	0.323
Watch, Clock and Optical Instruments Manuf.	0.575
Motor Vehicle Parts and Accessories Manuf.	0.880
Transportation Equipment Manuf.	0.909
All Other Miscellaneous Manuf.	-0.004
Metal and Non-Metal Scrap Recyclig	0.561
Production and Distribution of Electrical, Gas and Vapor Energy	1.010
Treatment and Distribution of Water	0.683
Construction Services & Installation Works	-0.274
Retail & Wholesale of Motor Vehicles & Parts	-0.093
Wholesale Trade in Goods and General Merchandise	0.203
Retail and Repair of Apparels, Food products	-0.296
Hotel Estab. & Food Services	-0.416
Rail and Road Passengers transportation	0.019
Maritime Shipping and Transport	0.165
Air and Space Tranport	0.749
Organization of Freight Transport	0.411
Telecommunications and Mail Activities	0.424
Banks, Investment Funds, Financial Intermed.	0.838

Insurances and Pension	0.697
Auxiliary Activities of Financial Intermediation	0.555
Real Estate	0.236
Transportation, Machinery, Equipment Rental	0.359
Software and Computer Development, and Other Computer Activities	0.380
R&D (Physical, Natural, Social and Human Sciences)	1.181
Service Activities to Businesses	-0.168
Public Admin., Defense, Justice & Foreign Relat.	0.035
Education and Teaching Activities	0.288
Health Related Activities and Social Services	0.043
Urban Cleaning and Sewerage	0.161
Professional, Political Organizations, Unions	0.232
Performing, Arts, Sports and Leisure Activities	0.238
Personal Service Activities	-0.320
Domestic Services	1.501
International Bodies and Other Extraterritorial Institutions	2.754

Table A.2.4: Average Fixed Effect Estimates by Microregion

Microregion Name	Rank	Average Fixed Effects
Rosario Oeste	Top1	1.205
Pio IX	Top2	0.913
Alto Solimoes	Top3	0.690
Macae	Top4	0.636
Brasilia	Top5	0.505
Boa Vista	Top6	0.425
Baia da Ilha Grande	Top7	0.422
Mazagao	Top8	0.388
Coari	Top9	0.387

Sao Jeronimo	Top10	0.364
Sao Jose dos Campos	Top11	0.335
Jundiai	Top12	0.331
Camaqua	Top13	.292
Sao Paulo	Top14	0.288
Japaratuba	Top15	0.286
Guarulhos	Top16	0.283
Campinas	Top17	0.276
Parauapebas	Top18	0.273
Macapa	Top19	0.269
Santos	Top20	0.261
Caxias do Sul	Top21	0.259
Rio Claro	Top22	.229
Tarauaca	Top23	0.216
Sorocaba	Top24	0.210
Joinville	Top25	0.202
Serra do Pereiro	Bottom25	-0.762
Sape	Bottom24	-0.763
Esperanca	Bottom23	-0.778
Serrinha	Bottom22	-0.783
Serido Oriental	Bottom21	-0.784
Juazeiro	Bottom20	-0.788
Serrana dos Quilombos	Bottom19	-0.802
Alfenas	Bottom18	-0.813
Baixo Jaguaribe	Bottom17	-0.813
Oliveira	Bottom16	-0.814
Furos de Breves	Bottom15	-0.817
Serido Ocidental Paraibano	Bottom14	-0.828
Itapemirim	Bottom13	-0.832

Chapada do Araripe	Bottom12	-0.857
Faxinal	Bottom11	-0.859
Ipu	Bottom10	-0.860
Litoral Nordeste	Bottom9	-0.951
Caririacu	Bottom8	-0.961
Patrocinio	Bottom7	-1.004
Brejo Santo	Bottom6	-1.066
Medio Curu	Bottom5	-1.071
Montanha	Bottom4	-1.135
Baixo Parnaiba Piauiense	Bottom3	-1.164
Coreau	Bottom2	-1.189
Lavras da Mangabeira	Bottom1	-1.299

Population By Sector, Microregion

Table A.2.5: RAIS: Top / Bottom 10 Microregions by Population Size

Rank	Name	Unit number	Total Employees	Share Employees
Top1	Sao Paulo	35061	2.835e+06	0.139
Top2	Rio de Janeiro	33018	1.685e+06	0.0825
Top3	Belo Horizonte	31030	937288	0.0459
Top4	Porto Alegre	43026	667190	0.0327
Top5	Curitiba	41037	660056	0.0323
Top6	Salvador	29021	550185	0.0270
Top7	Brasilia	53001	488386	0.0239
Top8	Campinas	35032	474091	0.0232
Top9	Recife	26017	455097	0.0223
Top10	Fortaleza	23016	447396	0.0219
Bottom10	Baixa Verde	24013	348.2	1.71e-05
Bottom9	Serrana do Sertao Alagoano	27001	315.8	1.55e-05
Bottom8	Baixo Parnaiba Maranhense	21013	309.3	1.52e-05
Bottom7	Caracarai	14003	287.9	1.41e-05
Bottom6	Mazagao	16004	276.3	1.35e-05
Bottom5	Alvorada D'Oeste	11005	274.2	1.34e-05
Bottom4	Pio IX	22014	271.4	1.33e-05
Bottom3	Rosario Oeste	51016	269	1.32e-05
Bottom2	Camaqua	43028	229.2	1.12e-05
Bottom1	Tabuleiro	42017	218.9	1.07e-05

Table A.2.6: Census: Top / Bottom 10 Microregions by Population Size

Rank	Name	Unit number	Total Employees
Top1	Sao Paulo	35061	807643
Top2	Rio de Janeiro	33018	692180
Top3	Belo Horizonte	31030	266329
Top4	Porto Alegre	43026	217302
Top5	Recife	26017	178077
Top6	Salvador	29021	172621
Top7	Curitiba	41037	165188
Top8	Fortaleza	23016	159207
Top9	Campinas	35032	140114
Top10	Brasilia	53001	111041
Bottom10	Caracarai	14003	2531
Bottom9	Almeirim	15003	2503
Bottom8	Paranatinga	51008	2386
Bottom7	Meruoca	23004	2109
Bottom6	Boca do Acre	13011	1957
Bottom5	Sena Madureira	12003	1954
Bottom4	Oiapoque	16001	1743
Bottom3	Japura	13002	1533
Bottom2	Amapa	16002	1409
Bottom1	Fernando de Noronha	26019	235

Table A.2.7: Distribution of employees by Sector (mean over years)

Name Sector	ID sector	Nb. Employees	Share Employees

Public Admin., Defense, Justice & Foreign Relat.	75	3.456e+06	0.169
Retail Trade, Repair of Apparels, Food products	52	2.597e+06	0.127
Service Activities Mainly provided to Businesses	74	1.953e+06	0.0957
Construction Services & Installation Works	45	1.246e+06	0.0611
Agriculture - Animal and Vegetable Products	1	935168	0.0458
Rail and Road Passengers transportation	60	903947	0.0443
Food Processing and Manuf.	15	816807	0.0400
Health Related Activities and Social Services	85	774900	0.0380
Wholesale Trade in Goods and General Merchandise	51	680757	0.0334
Hotel Estab. & Food Services	55	648124	0.0318
Education and Teaching Activities	80	566547	0.0278
Retail & Wholesale of Motor Vehicles & Parts	50	562975	0.0276
Real Estate	70	493791	0.0242
Professional, Political Organizations, Unions	91	350689	0.0172
Banks, Investment Funds, Financial Intermed.	65	301245	0.0148
Apparel Accessories Manuf.	18	301084	0.0148
Household Appliance & Equipments Manuf.	29	270184	0.0132
Metal Product Manuf.	28	260074	0.0127
Motor Vehicle Parts and Accessories Manuf.	34	245906	0.0121
Miscellaneous Chemical Product Manuf.	24	227025	0.0111
Plastics Product Manuf.	25	216485	0.0106
Nonmetallic Mineral Product Manuf.	26	182998	0.00897
Organization of Freight Transport	63	180536	0.00885
Textile Products Manuf.	17	174488	0.00855
Leather Products Manuf.	19	169791	0.00832
Telecommunications and Mail Activities	64	161179	0.00790
All Other Miscellaneous Manuf.	36	159000	0.00779
Nonferrous Metal Foundries	27	149612	0.00733
Software and Computer Dvpment, Consulting	72	144000	0.00706

Performing, Arts, Sports and Leisure Activities	92	126086	0.00618
Medias Publishing, Printing and Reproducing	22	111668	0.00547
Wood Products Manuf.	20	108103	0.00530
Electrical Machinery & Supplies Manuf.	31	93676	0.00459
Pulp, Paper and Paperboard Products Manuf.	21	78629	0.00385
Urban Cleaning and Sewerage	90	71182	0.00349
Personal Service Activities	93	61831	0.00303
Production and Distribution of Energy	40	60875	0.00298
Treatment and Distribution of Water	41	58213	0.00285
Transportation, Machinery, Equipment Rental	71	53717	0.00263
Insurances and Pension	66	52597	0.00258
Silviculture and Forest exploration	2	46656	0.00229
Transportation Equipment Manuf.	35	45505	0.00223
Audio and Video Equipment Manuf.	32	43866	0.00215
Coal & Nuclear Combustibles Processing	23	43033	0.00211
Air and Space Tranport	62	27250	0.00134
Petroleum and Gas Extraction	11	26744	0.00131
Auxiliary Activities of Financial Intermediation	67	26564	0.00130
Metal and radioactive Ore Mining	13	25856	0.00127
Watch, Clock and Optical Instruments Manuf.	33	22798	0.00112
Nonmetallic Mineral Mining	14	20884	0.00102
Office Machinery & Computer Manuf.	30	16746	0.000821
Research & Dvpt	73	14353	0.000704
Tobacco Product Manuf.	16	11227	0.000550
Maritime Shipping and Transport	61	8800	0.000431
Fishery	5	5980	0.000293
Coal Mining	10	2574	0.000126
Metal and Non-Metal Scrap Recyclig	37	2460	0.000121
Intl. Bodies & Extraterritorial Instit.	99	1145	5.61e-05

Domestic Services	95	441.5	2.16e-05

Transitions Between Microregions

For each demographic type - year - sector, we computed the number of individuals who moved across microregions. We created a dataset of transitions to record both the frequency of transitions from or towards a given microregion by year, sector and demographic type, as well as the income changes associated to the transitions.

The dataset contains 309,261,026 observations over years 2000-2014. For each year, demographic type, sector (2 digits), microregion of origin, microregion of destination, it records the number of individuals in the origin, the number of individuals who moved from the origin to the destination, the mean income in the origin, the mean income in the destination and the mean income of the individuals who moved.

Table A.2.8: Mean Population and Income over Years, Transitions

Variable	Mean	Std. dev.	Min	Max	p10	p50	p90
Nb. indiv. origin	2,151	6,853	1	155,120	2	410	4,152
Nb. indiv. moved	484	3,630	1	110,819	2	31	406
Mean income origin	7,175	13,130	0	1,030,152	692	3,536	15,880
Mean income dest.	9,015	14,541	51	467,627	1,966	4,821	18,685
Mean income movers	3,525	8,818	0	1,030,152	553	1,489	7,135

Table A.2.9: Top/ Bottom 10 most attractive origin microregions

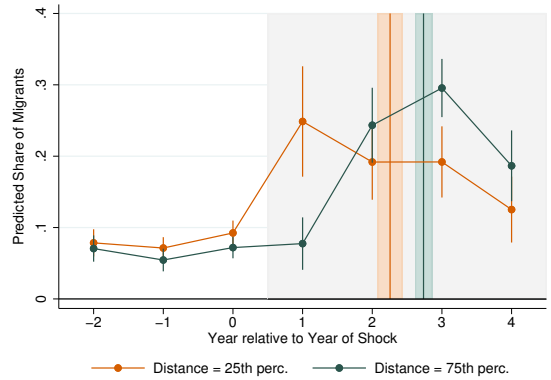
Microregion	Rank	Nb. Movers 1000s/year	Mean Inc. Origin	Mean Inc. Dest.	Mean Inc. Movers
Sao Paulo	Top1	1903	14566	9263	4908
Rio de Janeiro	Top2	777.4	15810	10522	7182
Belo Horizonte	Top3	735.7	11143	9416	4331
Osasco	Top4	610.5	8824	9440	3517
Curitiba	Top5	495.1	11201	9562	4404
Salvador	Top6	489.7	14229	11183	6375
Campinas	Top7	447.7	11917	9457	4719
Porto Alegre	Top8	429.9	14156	10614	5399
Brasilia	Top9	355.1	21812	11788	8928
Recife	Top10	332.1	11953	10553	5521
Batalha	Bottom10	3.113	7459	12828	10643
Medio Curu	Bottom9	3.106	6581	18015	7281
Jeremoabo	Bottom8	2.900	4121	8705	3985
Livramento do Brumado	Bottom7	2.875	4360	10965	2029
Santana do Ipanema	Bottom6	2.590	5668	12889	7043
Alvorada D'Oeste	Bottom5	2.444	3455	11121	4612
Boquira	Bottom4	2.269	3825	12547	4848
Meruoca	Bottom3	2.243	5783	7568	2516
Traipu	Bottom2	2.211	6239	16605	4011
Baixo Parnaiba Maranhense	Bottom1	2.062	2384	5562	4344

Table A.2.10: Top/ Bottom 10 most attractive sectors

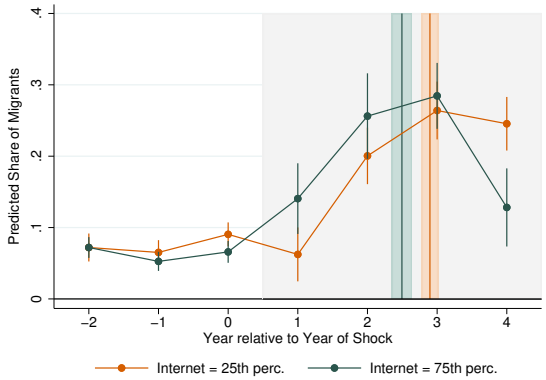
Rank	Sector	Nb. Movers 1000s/year	Mean Inc. Origin	Mean Inc. Dest.
Top1	Public Admin., Defense, Justice	1136	22082	19174
Top2	Retail Trade and Food products	1026	3726	3637
Top3	Construction Services & Installation Works	768.7	3171	3693
Top4	Service Activities Provided to Businesses	754.2	3526	4558
Top5	Agriculture - Animal & Vegetable Products	506.2	3907	4243
Top6	Health Related Activities & Social Services	442.1	7808	9648
Top7	Rail and Road Passengers transportation	336.5	4694	5429
Top8	Retail & Wholesale of Motor Vehicles	330.1	4886	5460
Top9	Wholesale Trade in Goods	323.2	6192	7395
Top10	Food Processing and Manuf.	320.6	6638	7067
Bottom10	Fishery	27.97	1005	2560
Bottom9	Domestic Services	27	2256	8701
Bottom8	Office Machinery & Computer Manuf.	26.99	5376	10509
Bottom7	Watch, Clock & Optical Manuf.	21.21	5504	14191
Bottom6	Tobacco Product Manuf.	12.82	7678	28579
Bottom5	Research and Development	12.12	19238	61006
Bottom4	Intl Bodies & Extraterritorial Instit.	10.56	26401	49175
Bottom3	Maritime Shipping and Transport	9.654	5984	17660
Bottom2	Metal and Non-Metal Scrap Recyclig	6.440	1955	5201
Bottom1	Coal Mining	5.556	1591	9554

A.2.2 Additional Suggestive Evidence on Information Frictions

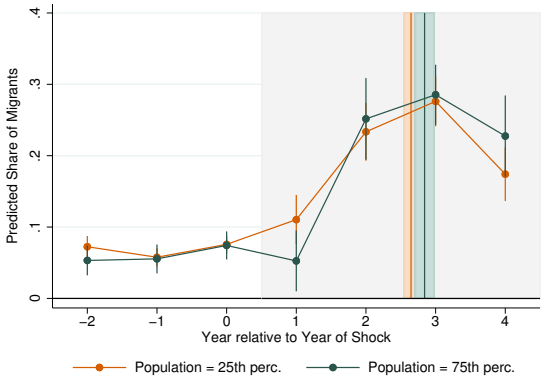
Figure A.2.1: Heterogeneity in Migration Responses: Labor Market Shock in Suape



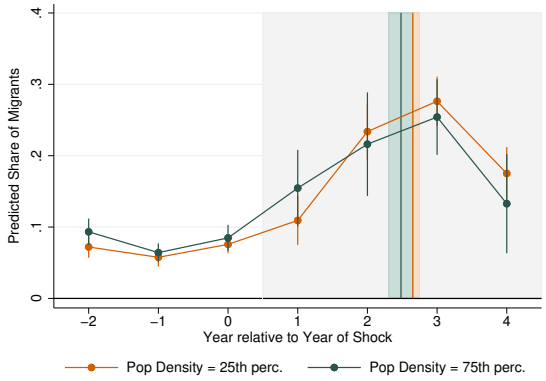
(a) Distance



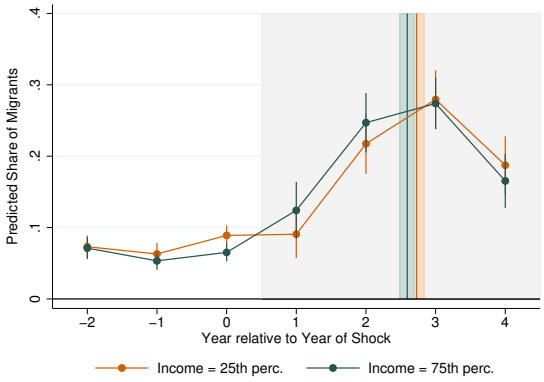
(b) Internet



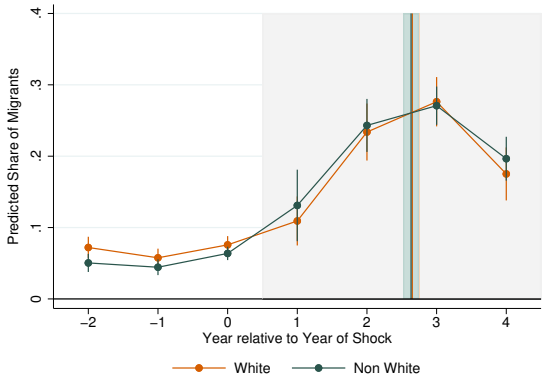
(c) Population



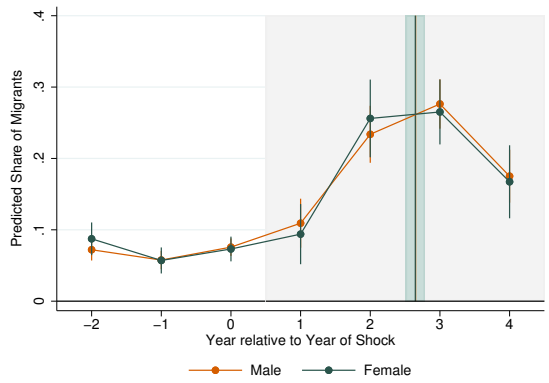
(d) Population Density



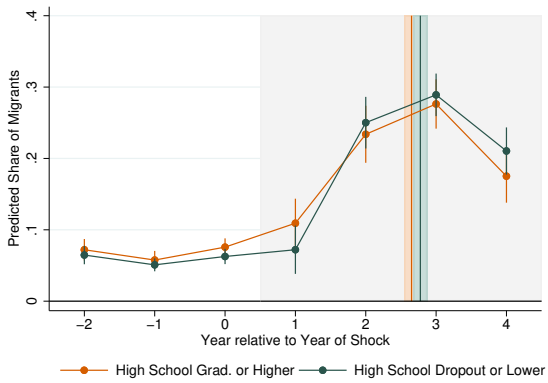
(e) Average Income



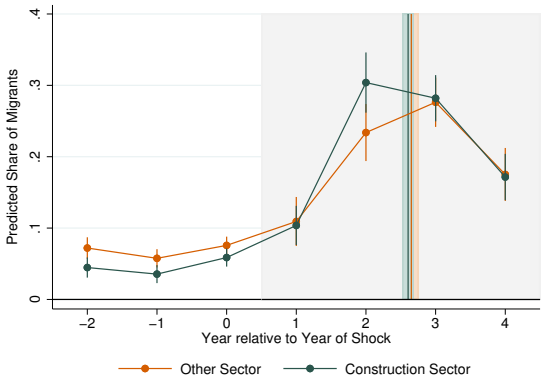
(f) Race



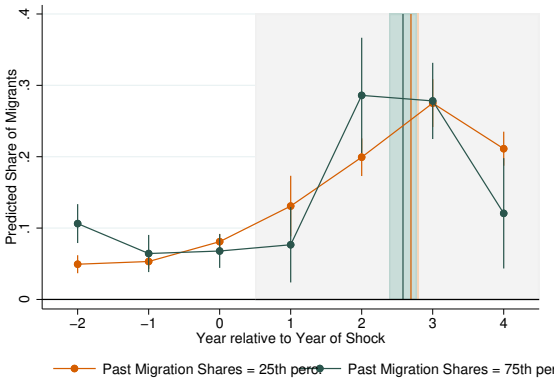
(g) Gender



(h) Education

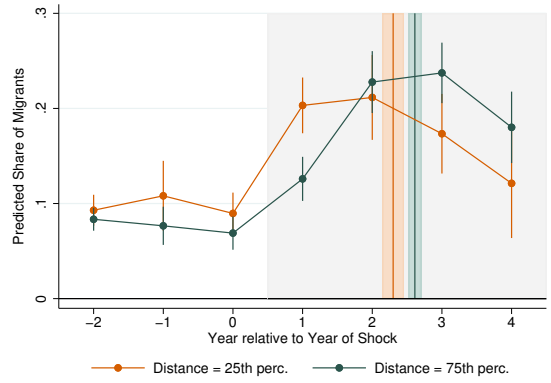


(i) Construction Sector

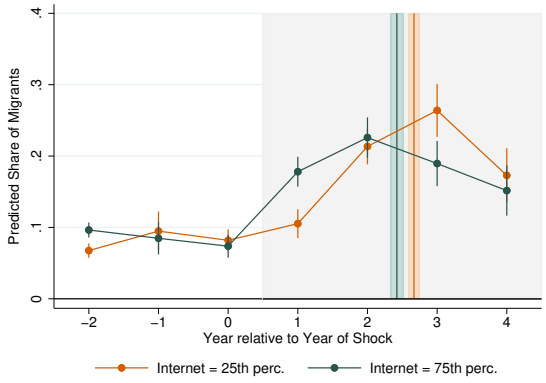


(j) Past Migration Flows

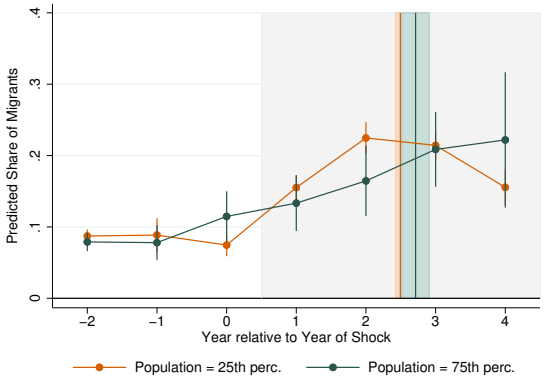
Figure A.2.2: Heterogeneity in Migration Responses: Labor Market Shock in Porto Velho



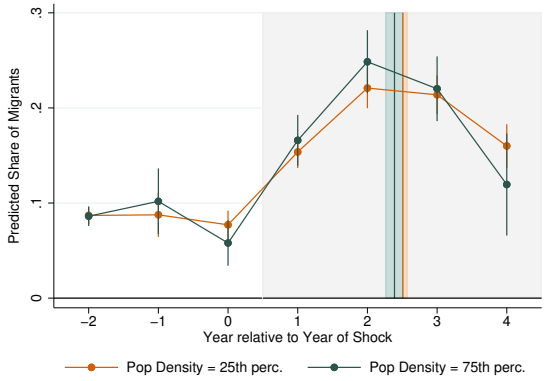
(a) Distance



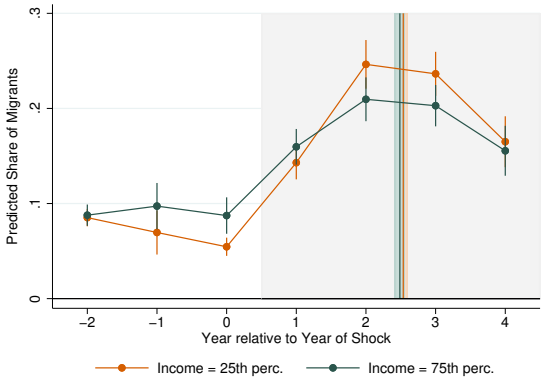
(b) Internet



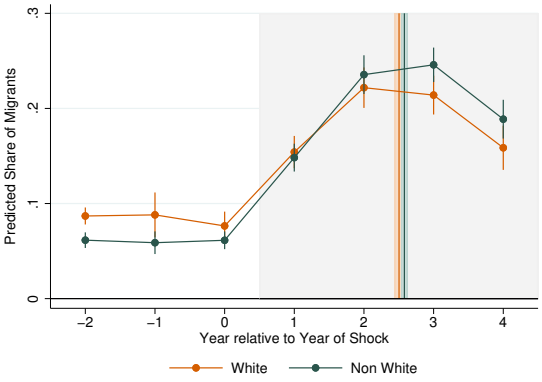
(c) Population



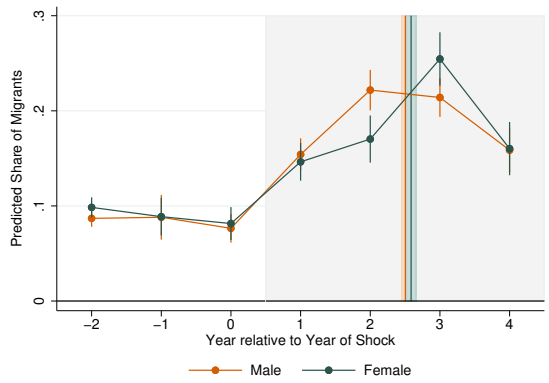
(d) Population Density



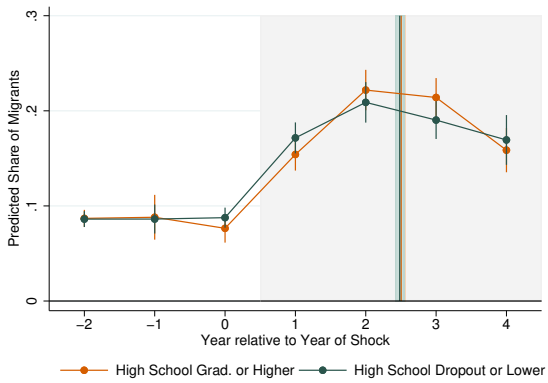
(e) Average Income



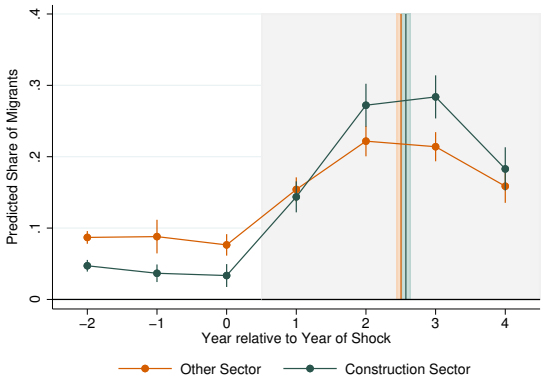
(f) Race



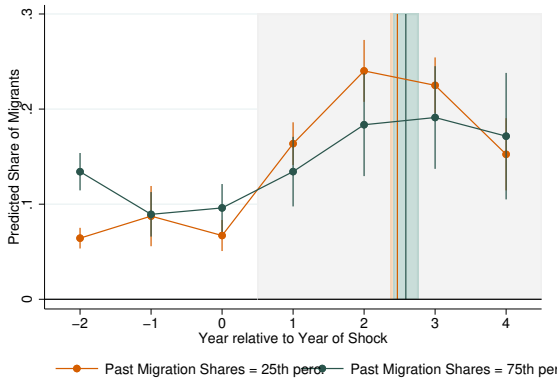
(g) Gender



(h) Education

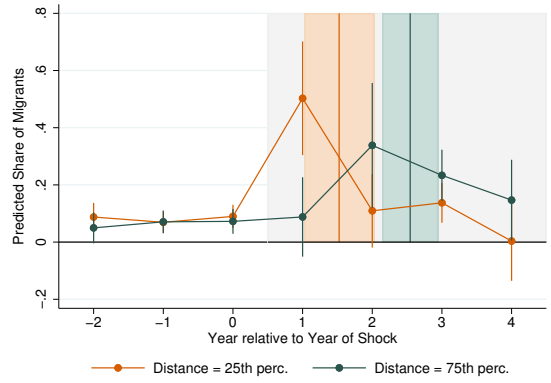


(i) Construction Sector

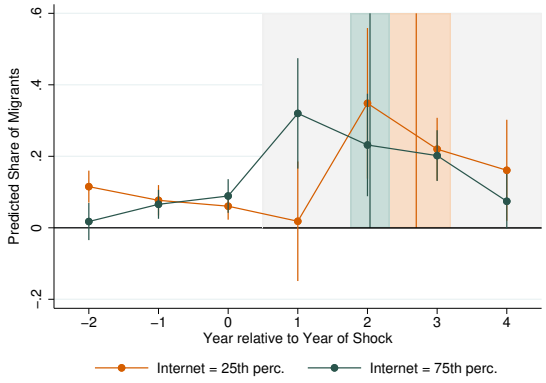


(j) Past Migration Flows

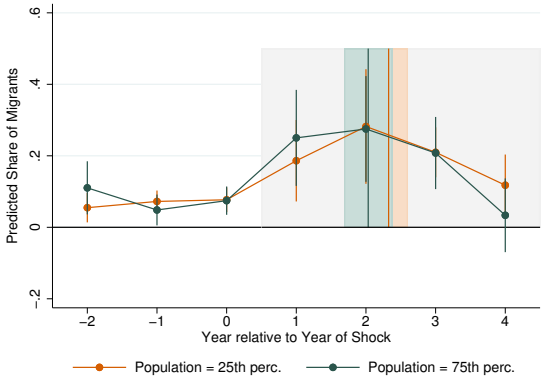
Figure A.2.3: Heterogeneity in Migration Responses: Labor Market Shock in Medio Mearim



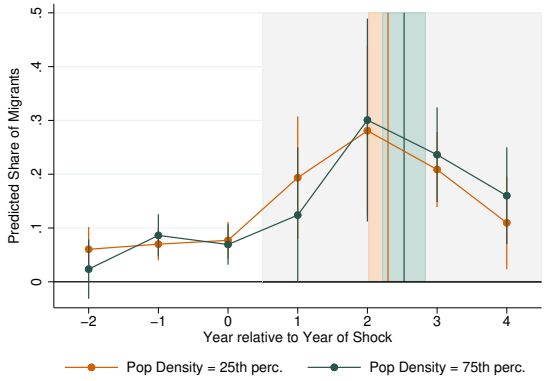
(a) Distance



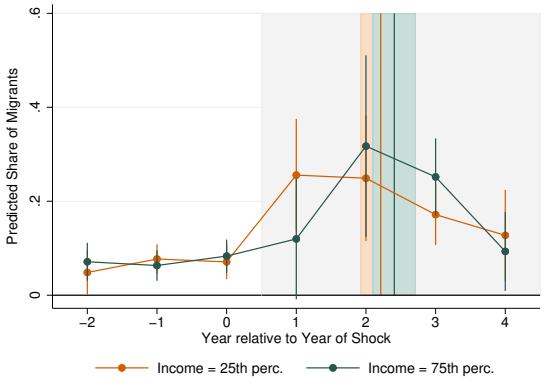
(b) Internet



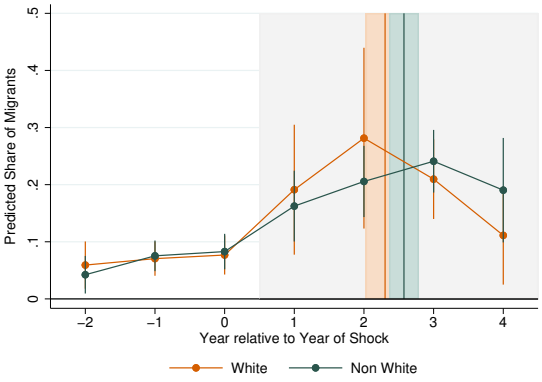
(c) Population



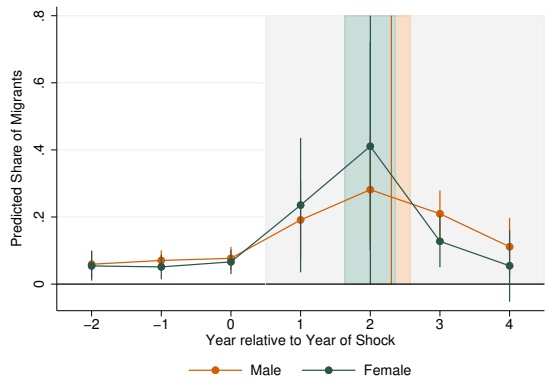
(d) Population Density



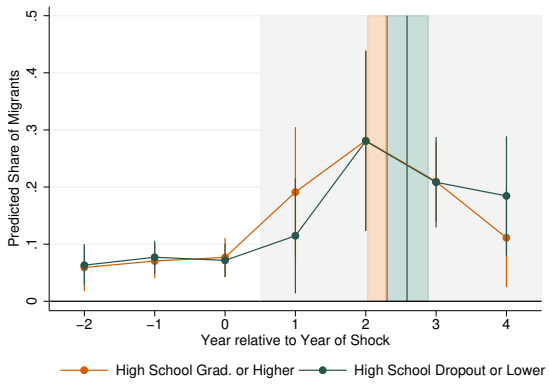
(e) Average Income



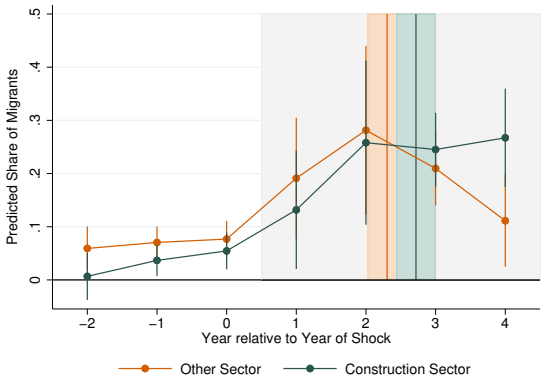
(f) Race



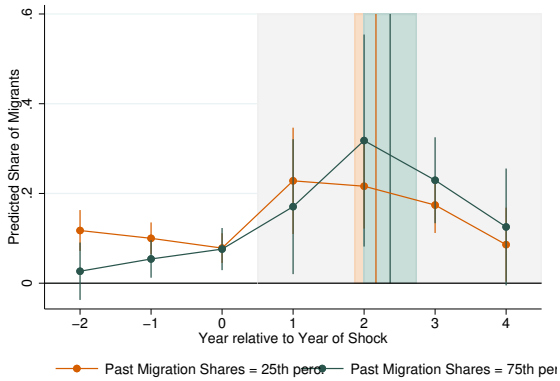
(g) Gender



(h) Education

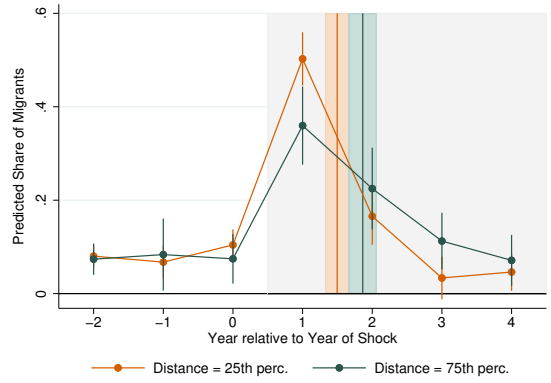


(i) Construction Sector

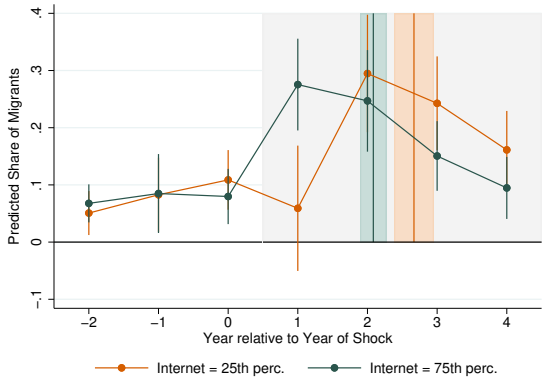


(j) Past Migration Flows

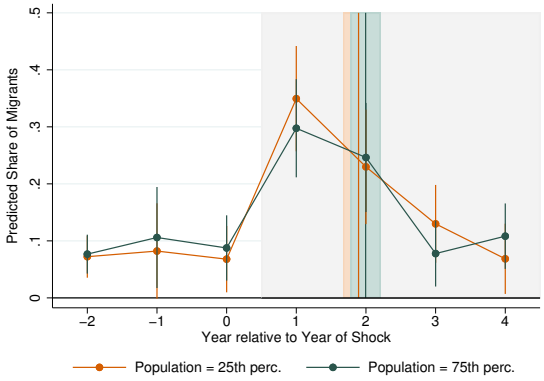
Figure A.2.4: Heterogeneity in Migration Responses: Labor Market Shock in Quirinópolis



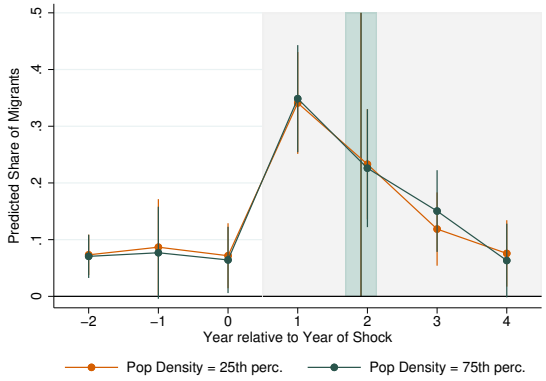
(a) Distance



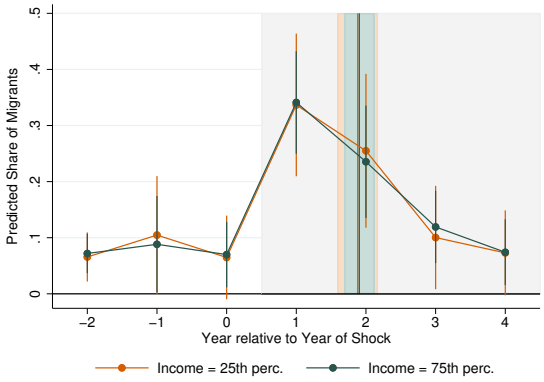
(b) Internet



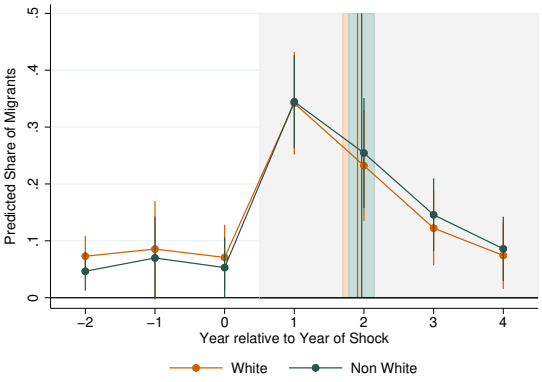
(c) Population



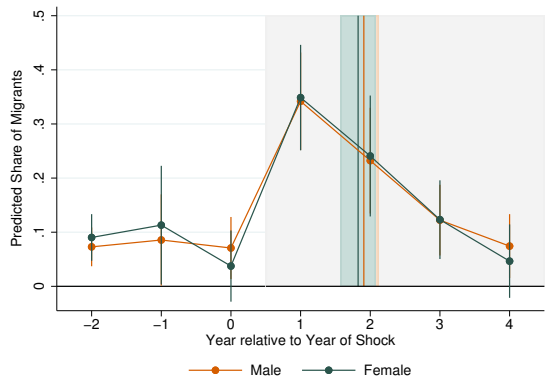
(d) Population Density



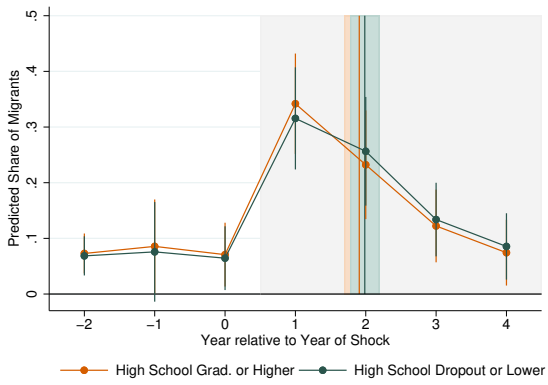
(e) Average Income



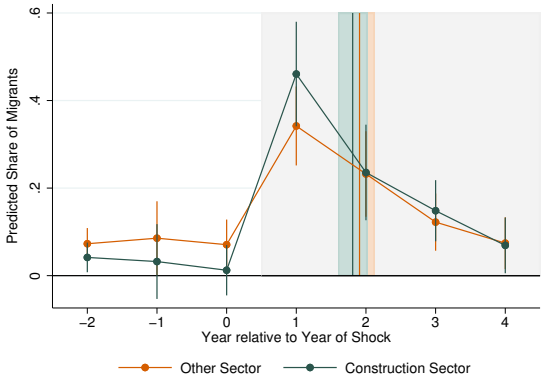
(f) Race



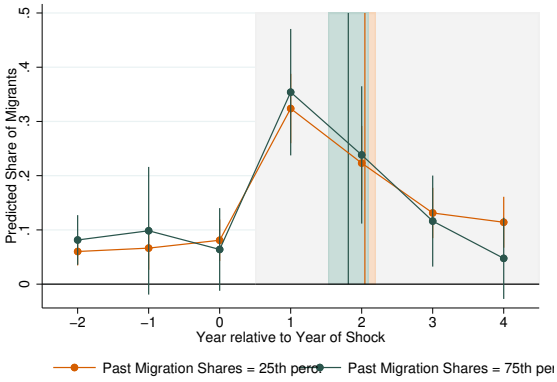
(g) Gender



(h) Education



(i) Construction Sector



(j) Past Migration Flows

A.3 Additional Results on Testing for the Content of Information Sets

A.3.1 Derivation of the Overidentification Test

Let's assume an estimating equation

$$\begin{aligned} & y(l, l_L(g), n_L(g), d(g), t) \\ = \; & \chi(l, l_L(g), d(g), t) + \alpha\big(w(l, n_L(g), d(g), t) - w(l_L(g), n_L(g), d(g), t)\big) \\ & + \varepsilon(l, l_L(g), n_L(g), d(g), t), \end{aligned} \tag{A.3.1}$$

and

$$\mathbb{E}[\varepsilon(l, l_L(g), n_L(g), d(g), t) | \mathcal{J}(l_L(g), n_L(g), d(g), t)] = 0,$$

where

$$\mathcal{J}(l_L(g), n_L(g), d(g), t) \equiv \{\mathcal{J}(l, l_L(g), n_L(g), d(g), t)\}_{l=1}^{L}.$$

To understand these expressions and how they map to our expressions, it is important to bear in mind that a group g corresponds to a group of agents sharing the same demographic group, the same lagged sector, and the same lagged location; i.e. if $g = g'$, then $(d', n'_L, l'_L) = (d, n_L, l_L)$. Consequently, indicating that a variable that varies with the vector $(l_L(g), n_L(g), d(g))$ is equivalent to indicating that it varies by group g.

The instrumental variables we would like to test take the following form:

$$\begin{aligned} & Z_{n,r,s,k}(l, l_L(g), n_L(g), d(g), t) \\ = \; & n\big(w(r(l), n_L(g), d(g), t-s) - w(r(l_L(g)), n_L(g), d(g), t-s)\big) \\ & \times \mathbb{1}\{(l, l_L(g), d(g), t) \in k\}, \end{aligned} \tag{A.3.2}$$

where $r(l)$ is a region to which microregion l belongs (i.e. $r(l)$ can either be the microregion l itself of the mesoregion to which region l belongs), $n(\cdot)$ is a function (i.e., it can either be an indicator function taking value 1 if $w(r(l), n_L(g), d(g), t-s) - w(r(l_L(g)), n_L(g), d(g), t-s)$ is positive or the exact difference $w(r(l), n_L(g), d(g), t-s) - w(r(l_L(g)), n_L(g), d(g), t-s)$ itself), s indicates the lag at which wages are measured, and k is a subset of possible values of the vector $(l, l_L(g), d(g), t)$.

Our baseline instrument corresponds to

$$\begin{aligned}
&\mathcal{Z}(l, l_L(g), n_L(g), d(g), t) \\
&= sign\big(w(M(l), n_L(g), d(g), t-1) - w(M(l_L(g)), n_L(g), d(g), t-1)\big) \\
&\times \mathbb{1}\{(l, l_L(g), d(g), t) \in \text{All}\} \\
&= sign\big(w(M(l), n_L(g), d(g), t-1) - w(M(l_L(g)), n_L(g), d(g), t-1)\big) \\
&= \mathbb{1}\big\{w(M(l), n_L(g), d(g), t-1) - w(M(l_L(g)), n_L(g), d(g), t-1) \geq 0\big\} \\
&- \mathbb{1}\big\{w(M(l), n_L(g), d(g), t-1) - w(M(l_L(g)), n_L(g), d(g), t-1) < 0\big\}.
\end{aligned}$$

where $M(l)$ is the mesoregion in which microregion l is located. Thus, we are imposing the assumption that

$$\mathbb{E}[\varepsilon(l, l_L(g), n_L(g), d(g), t) | \mathcal{Z}(l, l_L(g), n_L(g), d(g), t)] = 0. \tag{A.3.3}$$

For any alternative instruments $Z_{(1)}(l, l_L(g), n_L(g), d(g), t)$ and $Z_{(2)}(l, l_L(g), n_L(g), d(g), t)$, where these instruments correspond to that in equation (A.3.3) for specific choices of n, r, s and k other than those implied by $\mathcal{Z}(l, l_L(g), n_L(g), d(g), t)$, we would like to test that, *conditional on the mean-independence condition in equation* (A.3.3) *holding*, the following mean-independence condition also holds

$$\begin{aligned}
&\mathbb{E}\Big[\varepsilon(l, l_L(g), n_L(g), d(g), t)\Big| \\
&\mathcal{Z}(l, l_L(g), n_L(g), d(g), t), Z_{(1)}(l, l_L(g), n_L(g), d(g), t), Z_{(2)}(l, l_L(g), n_L(g), d(g), t)\Big] = 0.
\end{aligned} \tag{A.3.4}$$

At this point, we are not testing the mean independence restriction in equation (A.3.4) but we are using a χ^2 test of over-identifying restrictions to testing instead one implication of it; i.e.,

$$\mathbb{E}\Big[\varepsilon(l, l_L(g), n_L(g), d(g), t)\Big|\mathcal{Z}(l, l_L(g), n_L(g), d(g), t)\Big] = 0, \tag{A.3.5a}$$

$$\mathbb{E}\Big[\varepsilon(l, l_L(g), n_L(g), d(g), t)\Big|Z_{(1)}(l, l_L(g), n_L(g), d(g), t)\Big] = 0, \tag{A.3.5b}$$

$$\mathbb{E}\Big[\varepsilon(l, l_L(g), n_L(g), d(g), t)\Big|Z_{(2)}(l, l_L(g), n_L(g), d(g), t)\Big] = 0. \tag{A.3.5c}$$

Importantly, we are not currently testing whether equations (A.3.5b) and (A.3.5c) hold conditional on equation (A.3.5a) holding (what is what we should be doing if we truly want to test equation (A.3.4) conditional on the mean-independence condition in equation (A.3.3) holding), but rather whether the three moment conditions in equation (A.3.5) jointly hold.

To modify our testing procedure to properly condition on equation (A.3.5a) holding is not complicated: the exact formula for the test is in page 193 of Arellano (2003), under the label of *Incremental Sargan Tests;* and in pages 218-221 of Hayashi (2000). The exposition in Hayashi (2000) is particularly convenient for us, as the moments in this case are linear in the parameters.

A.3.2 Implementation of the Overidentification Tests

"Within" Expression

We turn to the "within" version of Equation (A.3.2) by subtracting the average of each variable within each region, past region, demographic group and year cell. Denoting by $\bar{x}(l, l_L(g), d(g), t)$ the average of x within the cell $(l, l_L(g), d(g), t)$, we get

$$y(l, l_L(g), n_L(g), d(g), t) - \bar{y}(l, l_L(g), d(g), t) = \alpha\big(\big(w(l, n_L(g), d(g), t) - w(l_L(g), n_L(g), d(g), t)\big) \tag{A.3.6}$$

$$- \big(\bar{w}(l, d(g), t) - \bar{w}(l_L(g), d(g), t)\big) \tag{A.3.7}$$

$$+ \varepsilon(l, l_L(g), n_L(g), d(g), t) - \bar{\varepsilon}(l, l_L(g), d(g), t), \tag{A.3.8}$$

and the error is mean-independent of the information set shared by workers in the same origin region, sector, demographic group and year,

$$\mathbb{E}[\varepsilon(l, l_L(g), n_L(g), d(g), t) | \mathcal{J}(l_L(g), n_L(g), d(g), t)] = 0,$$

where

$$\mathcal{J}(l_L(g), n_L(g), d(g), t) \equiv \bigcap_{l=1}^{L} \{\mathcal{J}(l, l_L(g), n_L(g), d(g), t)\}.$$

The average error, however, $\bar{\varepsilon}(l, l_L(g), d(g), t)$ is only mean-independent of the information set shared by workers in the same origin region, demographic group and year, across all sectors, so

$$\mathbb{E}[\bar{\varepsilon}(l, l_L(g), n_L(g), d(g), t) | \mathcal{J}(l_L(g), d(g), t)] = 0,$$

where

$$\mathcal{J}(l_L(g), d(g), t) \equiv \bigcap_{l=1}^{L} \bigcap_{n=1}^{N} \{\mathcal{J}(l, l_L(g), n_L(g), d(g), t)\}.$$

We are not imposing that individuals that share an origin region, demographic group and year but are in different sectors have the same information set. They may be different, but we will not be able to test for heterogeneity of information sets across sectors since we can only use instruments that belong to the information sets of people in all sectors.

Denoting $\tilde{x} = x - \bar{x}$, we can rewrite (A.3.8) as

$$\tilde{y}(l, l_L(g), n_L(g), d(g), t) = \alpha\big(\tilde{w}(l, n_L(g), d(g), t) - \tilde{w}(l_L(g), n_L(g), d(g), t)\big) \tag{A.3.9}$$

$$+ \tilde{\varepsilon}(l, l_L(g), n_L(g), d(g), t), \tag{A.3.10}$$

and

$$\mathbb{E}[\tilde{\varepsilon}(l, l_L(g), n_L(g), d(g), t)|\mathcal{J}(l_L(g), d(g), t)] = 0,$$

In the following, we omit the indices denoting regions, sectors, demographic groups or years, and turn to a matrix notation. Let us write

$$y \equiv \tilde{y}(l, l_L(g), n_L(g), d(g), t),$$

$$X \equiv \tilde{w}(l, n_L(g), d(g), t) - \tilde{w}(l_L(g), n_L(g), d(g), t),$$

$$\varepsilon \equiv \tilde{\varepsilon}(l, l_L(g), n_L(g), d(g), t),$$

$$Z \equiv \mathcal{Z}(l, l_L(g), n_L(g), d(g), t)$$

so that (A.3.10) writes

$$y = \alpha X + \varepsilon. \tag{A.3.11}$$

J-Statistic with Two-Way Clustered Standard Errors

Given that the expectational errors about wages for any location may be correlated across people in different origins, sectors and years (unobserved productivity shocks that are location specific and potentially serially correlated), and given that our specification differences out two locations (one of them chosen to be the prior location of the agents, i.e. the origin), we decide to two-way cluster our standard errors at the origin and destination level.[1]

Denote by n the number of observations in the vector y, K the number of instrumental variables and L the number of regressors. Following Hayashi p. 206, the GMM estimates $\hat{\alpha}$ of α minimizes

$$J(\tilde{\alpha}, \hat{W}) = n g_n(\tilde{\alpha})' \hat{W} g_n(\tilde{\alpha}),$$

[1]Note that this clustering already accounts for the additional correlation across residuals brought about by the differencing of the mean at the destination-origin-demographic-year level.

where $\hat{W}$ is a $K \times K$ symmetric positive definite weighting matrix, and $g_n(\tilde{\alpha})$ is the sample moment of g, respectively defined as

$$g_n(\tilde{\alpha}) = \frac{1}{n}\sum_{i=1}^{n} Z_i(y_i - X_i'\tilde{\alpha}) = s_{zy} - S_{zx}\tilde{\alpha}, \qquad g \equiv \mathbb{E}[g_i] = \mathbb{E}[Z_i(y_i - X_i'\alpha)],$$

where s_{zy} and S_{zx} are the sample moments of $\sigma_{zy} \equiv \mathbb{E}[Z_i y_i]$ and $\Sigma_{zx} = \mathbb{E}[Z_i X_i']$.[2]

The unique solution is[3]

$$\hat{\alpha}(\hat{W}) = (S_{zx}'\hat{W}S_{zx})^{-1}S_{zx}\hat{W}s_{zy}.$$

Denote by S the variance-covariance matrix of the moments $S \equiv \mathbb{E}[g_i g_i']$. If the weighting matrix $\hat{W}$ is chosen optimally to be a consistent estimator $\hat{S}^{-1}$ of S^{-1}, so that $p\lim \hat{W} = S^{-1}$, then the minimized distance J is asymptotically chi-squared, i.e.[4]

$$J(\hat{\alpha}(\hat{S}^{-1}), \hat{S}^{-1}) \to_d \chi^2(K - L).$$

Under the assumptions of linearity, that (y_i, X_i) is stationary and ergodic,[5] and that there exists a fourth moment for the instrumental variables[6], a consistent estimator of S is[7]

$$\hat{S} = \frac{1}{n}\sum_{i=1}^{n} \hat{\varepsilon}_i^2 Z_i Z_i', \tag{A.3.12}$$

where $\hat{\varepsilon}_i \equiv y_i - X_i'\hat{\alpha}_{2SLS}$, i.e. the residuals obtained from estimating α by two-stage least squares, as prescribed by the two-step efficient GMM procedure.[8]

[2] The sample moments are defined p. 201.

[3] This is given p. 207.

[4] This is proposition 3.6 p. 217.

[5] These are Assumptions 2.1 and 2.2 p. 109. The definition of *stationary* is defined p. 98 and that of *ergodicity* is given p. 101.

[6] This is Assumption 2.6 p.123, namely that $\mathbb{E}[(Z_{ik}Z_{ij})^2]$ exists and is finite for all k and $j = (1, 2, \ldots, K)$.

[7] This is Proposition 2.4 p. 124. This is also mentioned in the GMM context p. 212.

[8] See p. 213. Note that the 2SLS is the GMM estimator with weighting matrix $\hat{W} = S_{zz}^{-1}$, where $S_{zz} = \frac{1}{n}\sum_{i=1}^{n} Z_i Z_i'$. This is obtained p.226.

However, in the context of clustered errors, the stationarity assumption is not satisfied, so that (A.3.12) is not a consistent estimate of S. Cameron and Miller note that a consistent estimate of S in the case of multi-way clustering is[9]

$$\hat{S}_{\text{multiway}} = \frac{1}{n}\sum_{i=n}\sum_{j=1} Z_i Z_j' \hat{\varepsilon}_i \hat{\varepsilon}_j \mathbb{1}[i, j \text{ share any cluster}]$$

For two-way clustering, this robust variance estimator is easy to implement.[10] First, obtain three different cluster-robust matrices for the estimator by one-way clustering in, respectively, the first dimension, the second dimension, and by the intersection of the first and second dimensions. Then add the first two variance matrices and, to account for double-counting, subtract the third. Thus

$$\hat{S}_{2\text{way}} = \hat{S}_1 + \hat{S}_2 - \hat{S}_{1\cap 2},$$

with

$$\hat{S}_c = \frac{1}{n}\sum_{i=n}\sum_{j=1} Z_i Z_j' \hat{\varepsilon}_i \hat{\varepsilon}_j \mathbb{1}[i, j \text{ share cluster } c], \qquad c \in \{1, 2, 1\cap 2\}.$$

Finally, the computation of $\hat{S}_c$ can be done simply by sorting observations by group. Denote by G_c the total number of groups for the chosen cluster level c, g one such group, and $(Z_g, \hat{\varepsilon}_g)$ the sub-vectors of $(Z, \hat{\varepsilon})$ of observations in group g. Then we can write $\hat{S}_c$ as

$$\hat{S}_c = \frac{1}{n}\sum_{g=1}^{G_c} Z_g' \hat{\varepsilon}_g \hat{\varepsilon}_g' Z_g$$

[9] Equation (20) p. 336 of Cameron and Miller (2014).

[10] As mentioned by Cameron and Miller (2014) p. 337, unlike the standard one-way cluster case, $\hat{S}_{2\text{way}}^{-1}$ is not guaranteed to be positive semi-definite, so it is possible that it may lead to negative variances and J-statistics. Cameron, Gelbach, and Miller (2011) present an eigendecomposition technique used in the time series HAC literature that zeroes out negative eigenvalues in $\hat{S}_{2\text{way}}^{-1}$ to produce a positive semi-definite variance matrix estimate.

One can then implement the GMM estimation with $\hat{W} = (\hat{S}_{2\text{way}})^{-1}$. Namely, the J statistic at the basis of the overidentification test writes

$$J = n g_n(\hat{\alpha})' (\hat{S}_{2\text{way}})^{-1} g_n(\hat{\alpha}), \tag{A.3.13}$$

Overidentification Test with Clustered Standard Errors

We saw that when $\hat{W}$ is chosen to be a consistent estimate of S^{-1}, the J statistic was asymptotically distributed as a χ^2. Since we want to compare the performance of different sets of instrumental variables, say Z_{ref} and Z_{test}, we can compare the J statistic coming from the GMM estimation of α, using alternatively Z_{ref} and Z_{test}. Intuitively, if we believe that the instruments Z_{ref} satisfy the exclusion restriction, then It can then be compared to the p-value of $\chi^2(K-L)$, i.e.

$$p_{95} = 1 - \Phi(J; K-L),$$

where $\Phi(\cdot; K-L)$ is the cdf of the χ^2 distribution with $K-L$ degrees of freedom. Note that according to Proposition 3.7 p. 220,

$$J - J_1 \to_d \chi^2(K - K_1),$$

A.3.3 Additional Tests of the Content of Workers' Information Sets

Table A.3.1: Testing Content of Information Set - Static Model

Regions tested	Variable tested	J-test statistic	J-test p-value
All	$Z_1(g,t)$	28.61	0.00
$1000 < dist(l_L(g), \check{l}(m))$	$Z_1(g,t)$	48.67	0.00
$500 < dist(l_L(g), \check{l}(m)) < 1000$	$Z_1(g,t)$	24.65	0.00
$200 < dist(l_L(g), \check{l}(m)) < 500$	$Z_1(g,t)$	17.13	0.00
$dist(l_L(g), \check{l}(m)) < 200$	$Z_1(g,t)$	7.10	0.01
$internet(l_L(g), t) < 5\%$	$Z_1(g,t)$	18.44	0.00
$5\% < internet(l_L(g), t) < 15\%$	$Z_1(g,t)$	17.54	0.00
$15\% < internet(l_L(g), t) < 30\%$	$Z_1(g,t)$	8.54	0.01
$30\% < internet(l_L(g), t)$	$Z_1(g,t)$	5.66	0.03
All	$Z_2(g,t)$	23.92	0.00
$1000 < dist(l_L(g), \check{l}(m))$	$Z_2(g,t)$	14.82	0.00
$500 < dist(l_L(g), \check{l}(m)) < 1000$	$Z_2(g,t)$	11.22	0.00
$200 < dist(l_L(g), \check{l}(m)) < 500$	$Z_2(g,t)$	5.57	0.02
$dist(l_L(g), \check{l}(m)) < 200$	$Z_2(g,t)$	7.08	0.01
$internet(l_L(g), t) < 5\%$	$Z_2(g,t)$	9.75	0.00
$5\% < internet(l_L(g), t) < 15\%$	$Z_2(g,t)$	3.38	0.07
$15\% < internet(l_L(g), t) < 30\%$	$Z_2(g,t)$	4.84	0.03
$30\% < internet(l_L(g), t)$	$Z_2(g,t)$	3.50	0.06
All	$Z_3(g,t)$	4.61	0.04
$1000 < dist(l_L(g), \check{l}(m))$	$Z_3(g,t)$	7.26	0.01
$500 < dist(l_L(g), \check{l}(m)) < 1000$	$Z_3(g,t)$	4.12	0.04
$200 < dist(l_L(g), \check{l}(m)) < 500$	$Z_3(g,t)$	1.80	0.18
$dist(l_L(g), \check{l}(m)) < 200$	$Z_3(g,t)$	0.08	0.78
$internet(l_L(g), t) < 5\%$	$Z_3(g,t)$	2.46	0.12
$5\% < internet(l_L(g), t) < 15\%$	$Z_3(g,t)$	1.68	0.20
$15\% < internet(l_L(g), t) < 30\%$	$Z_3(g,t)$	0.81	0.49
$30\% < internet(l_L(g), t)$	$Z_3(g,t)$	0.64	0.42

Note: the variable $dist(l_L(g), \check{l}(m))$ denotes the distance (in kilometers) between microregions $l_L(g)$ and $\check{l}(m)$. The variable $internet(l_L(g), t)$ denotes the number of broadband internet connections per capita in microregion $l_L(g)$ in period $t-1$.

Appendix B

Appendix for "Migration with Costly Information"

B.1 Appendix to Section 2.2

B.1.1 Proofs

Proof of Social Learning Rule

Following Molavi et al. (2018), I postulate that agents follow social learning rules that satisfy *imperfect recall*, according to which they treat the current beliefs of their neighbors as sufficient statistics for all the information available to them while ignoring how or why these opinions were formed. This is a formalization of the idea that real-world individuals do not fully account for the information buried in the entire past history of actions or the complex dynamics of beliefs over social networks. Agents take the current beliefs of their neighbors as sufficient statistics for all the information available to them while ignoring how or why those opinions were formed. Denoting by $\bar{\pi}_{kt}$ the belief resulting from information sharing, imperfect recall implies that $\bar{\pi}_{kt}$ is only a function of the current beliefs of all agents in k at t:

Assumption 3 (Imperfect Recall)**.** *$\bar{\pi}_{kt}$ is independent of $\pi_{\tau-1}$ for all k and all $\tau \leq t-1$.*

In order to obtain a simple unique characterization of the social learning rule, I follow Molavi et al. (2018) and impose three natural additional restrictions on how agents process their neighbors' information. The first one is that agents' social learning rules are *label neutral*, which means that relabeling the underlying states has no bearing on how agents process information. Second, I assume that individuals do not discard their neighbors' most recent observations by requiring their social learning rules to be increasing in their neighbors' last period beliefs, a property referred to as *monotonicity*. Third, I require agents' learning rules to satisfy *independence of irrelevant alternatives*: each agent treats her neighbors' beliefs about any subset of states as sufficient statistics for their collective information regarding those states. The formal representation is reported in Appendix B.1.1.

Molavi et al. (2018) show that, in conjunction with imperfect recall, these three restrictions lead to a unique representation of agents' social learning rules up to a set of constants: at any given time period, the log-likelihood ratios of all agents' beliefs are combined linearly,

weighted by their centrality in the network. Given the assumption that agents are all connected to each other locally, they have the same centrality and the beliefs of every agents are given the same weight.[1]

Proposition 5. *Information sharing leads to a log-linear learning rule. The beliefs held by people in k at the end of t after local information sharing are:*

$$\log \bar{\pi}_{kt}(\theta_t) = C_{kt} + \sum_j \sum_{s_t} L_{jkt|s_t} \log \pi_{jt|s_t}(\theta_t), \tag{B.1.1}$$

where $L_{jkt|s} = L_{jt-1}\bar{p}_{jkt}(\theta_t, L_{t-1}, \pi_{jt}, s_t)$ is the mass of agents from j in k who received signal s_t, and C_{kt} is a constant ensuring that $\int_\theta \bar{\pi}_{kt}(\theta)d\theta = 1$.

The proof is adapted from Theorem 1 in Molavi et al. (2018). I omit the time indices for brevity. Consider two arbitrary states $\theta \neq \hat{\theta}$ and an arbitrary profile of beliefs in each region $\pi = \{\pi_{j|s}\}_{s,j=1,\dots,J} \in \Delta\Theta^{J\times S}$. Let $\bar{\Theta} = \{\theta, \hat{\theta}\}$. Denote by $\bar{\pi}(\pi) = \{\bar{\pi}_k(\pi)\}_{k=1,\dots,J}$ the shared beliefs. By definition of conditional probability, for every region k:

$$\log \frac{\bar{\pi}_k(\pi)(\theta)}{\bar{\pi}_k(\pi)(\hat{\theta})} = \log \bar{\pi}_k \left(cond_{\bar{\Theta}}(\pi)\right)(\theta) - \log \bar{\pi}_k \left(cond_{\bar{\Theta}}(\pi)\right)(\hat{\theta}).$$

Note that $cond_{\bar{\Theta}}(\pi)$ depends on the belief profile π only through the collection of likelihood ratios $\{\pi_j(\theta)/\pi_j(\hat{\theta})\}$. Consequently, indexing all agents in k by $i \in [0,1]$, for any given region k, there exists a continuous function $g_k : \mathbb{R}^I \to \mathbb{R}$ such that:

$$\log \frac{\bar{\pi}_k(\pi)(\theta)}{\bar{\pi}_k(\pi)(\hat{\theta})} = g_k\left(\log \frac{\pi_0(\theta)}{\pi_0(\hat{\theta})}, \dots, \log \frac{\pi_1(\theta)}{\pi_1(\hat{\theta})}\right). \tag{B.1.2}$$

for all pairs of states $\theta \neq \hat{\theta}$ and all profiles of beliefs π. Furthermore, label neutrality guarantees that the function g_k is independent of θ and $\hat{\theta}$.

Now, consider three distinct states θ, $\hat{\theta}$ and $\tilde{\theta}$. Given that (B.1.2) has to be satisfied for any arbitrary pair of states, we have:

[1] As shown by Levy and Razin (2018), this log-linear rule can be obtained if agents treat their marginal information sources as conditionally independent.

$$
\begin{aligned}
& g_k\left(\log\frac{\pi_0(\theta)}{\pi_0(\hat{\theta})},\ldots,\log\frac{\pi_1(\theta)}{\pi_1(\hat{\theta})}\right)+g_k\left(\log\frac{\pi_0(\hat{\theta})}{\pi_0(\tilde{\theta})},\ldots,\log\frac{\pi_1(\hat{\theta})}{\pi_1(\tilde{\theta})}\right)\\
= \quad & \log\frac{\bar{\pi}_k(\pi)(\hat{\theta})}{\bar{\pi}_k(\pi)(\tilde{\theta})}+\log\frac{\bar{\pi}_k(\pi)(\hat{\theta})}{\bar{\pi}_k(\pi)(\tilde{\theta})}\\
= \quad & g_k\left(\log\frac{\pi_0(\theta)}{\pi_0(\tilde{\theta})},\ldots,\log\frac{\pi_1(\theta)}{\pi_1(\tilde{\theta})}\right)
\end{aligned}
$$

Since π was arbitrary, the above equation implies that for any arbitrary $x, y \in \mathbb{R}^I$, it must be the case that $g_k(x)+g_k(y)=g_k(x+y)$. This equation is nothing but Cauchy's functional equation, with linear functions as its single family of continuous solutions. Therefore, there exist constants a_{ik} such that $g_k(x)=\int_I a_{ik}x_i di$. Thus, using (B.1.2) one more time implies that

$$
\log\frac{\bar{\pi}_k(\pi)(\theta)}{\bar{\pi}_k(\pi)(\hat{\theta})}=\int_I a_{ik}\log\frac{\pi_i(\theta)}{\pi_i(\hat{\theta})}di, \quad \forall\theta,\hat{\theta}\in\Theta.
$$

Monotonicity implies that $a_{ik}>0$ for all $i \in I$. Since I assume that the weight that every agents place on each other individual's belief is $a_{ik}=1$, we can aggregate all beliefs that are identical and write:

$$
\log\frac{\bar{\pi}_k(\pi)(\theta)}{\bar{\pi}_k(\pi)(\hat{\theta})}=\sum_j\sum_s L_{jk|s}\log\frac{\pi_{jk|s}(\theta)}{\pi_{jk|s}(\hat{\theta})}. \tag{B.1.3}
$$

Proof of Lemma 1

Adapted from Lemma 1 in Steiner et al. (2017). The proof relies on the concavity of the entropy function. To prove the Lemma, I show that the discounted expected payoff from any strategy (σ, f) is equal to the value of the objective function in (2.15), given the choice rule generated by (σ, f).

Let (σ, f) be a strategy and p the choice rule generated by (σ, f), so that:

$$
p_{l_{t-1}l_t t}(\theta_t, L_{t-1}, \pi_t, \varepsilon_{it})=\Pr\left(\sigma_t(l_{t-1}, s_t, \pi_{l_{t-1}t}, L_{t-1}, \varepsilon_{it})=l_t|\theta_t, l_{t-1}, L_{t-1}, \pi_t, \varepsilon_{it}\right)
$$

First, let's show that the discounted expected payoff from p is at least as large as the one from (σ, f). By construction, (σ, f) and p give the same stream of expected gross payoffs. Remains to show that the stream of information costs associated with p is no larger than that associated with (σ, f).

Recall from (2.9) that in a given period t, the information cost associated with (σ, f) writes:

$$I_j\left(f|\pi_{jt}\right) = \lambda_j\left(H\left(\pi_{jt}\right) - \mathbb{E}_s\left[H\left(\pi_{jt|s}\right)\right]\right),$$

while the information cost associated with p writes:

$$I_j\left(p|\pi_{jt}\right) = \lambda_j\left(H\left(\pi_{jt}\right) - \sum_k q_{jkt} H\left(\pi_{jt|k}\right)\right),$$

where $q_{jkt} = \Pr\left(s_t = k|\pi_{jt}, p\right)$ is the ex-ante probability that p will send the signal $s_t = k$.

Note that l_t is measurable with respect to s_t:

$$\Pr(\theta_t|l_t) = \sum_{s_t} \Pr(\theta_t|s_t) \Pr(s_t|l_t),$$

Thus, $\Pr(\theta_t|l_t)$ is a convex combination of the distributions $\Pr(\theta_t|s_t)$, as s_t varies. By concavity of the entropy, this implies that

$$\mathbb{E}_s\left[H\left(\pi_{jt|s}\right)\right] \leq \sum_k q_{jkt} H\left(\pi_{jt|k}\right).$$

This shows that $I_j\left(p|\pi_{jt}\right) \leq I_j\left(f|\pi_{jt}\right)$, and hence that the discounted expected payoff from (σ, f) is no larger than the value of the objective function in (2.15), given the choice rule p generated by (σ, f).

Conversely, the discounted expected payoff from any strategy induced by a choice rule p is identical to the value of the objective function in (2.15) given p. Together, these two relationships imply the result.

Proof of Proposition 1

Consider the space of strategies induced by the mobility rule p mapping $(\theta_t, L_{t-1}, \pi_t, l_{t-1}, \varepsilon_{it})$ to an action l_t: $\Pi = \Delta(J)^{\Theta \times \mathcal{L} \times \Delta\Theta \times J \times \mathbb{R}^J}$.

The flows of payoffs, net of information costs, that are being maximized are:

$$u_{l_{t-1}l_t}(\theta_t, L_t, \varepsilon_{it}) = \theta_{l_t t} + A_{l_t} - \alpha \log L_{l_t t} + B_{l_t} - \kappa_{l_{t-1}l_t} + \nu \varepsilon_{l_t t}.$$

First, note that if $\alpha = 0$, and ε_{it} and θ_t were bounded above and below, then u would be uniformly bounded and hence continuous, and the space of strategies Π would be compact as a product of compact spaces by Tychonoff's theorem, so that an optimum to (2.15) would exist.

When $\alpha > 0$, u is still bounded from below, since the population in a given region is bounded by the total population, $L_{l_t t} \leq \bar{L}$. It is also easy to show that for any finite values of migration costs κ, productivities A, and amenities B, no equilibrium would feature $L_{l_t t} = 0$ since it would imply infinite payoffs in region l_t and the region would attract workers to increase $L_{l_t t}$.

We can therefore consider an auxiliary problem with bounded population, bounded productivities and bounded preference shocks characterized by $(b_L, b_\theta, b_\varepsilon)$ so that solves (2.15) with the additional constraint on the states $(L_{t-1}, \theta_t, \varepsilon_{it}) \in \mathcal{B}(b_L, b_\theta, b_\varepsilon)$, where:

$$\mathcal{B}(b_L, b_\theta, b_\varepsilon) = \{(L_{t-1}, \theta_t, \varepsilon_{it}) \mid L_{kt} > b_L,\ \theta_{kt} \in (-b_\theta, b_\theta),\ \varepsilon_{ikt} \in (-b_\varepsilon, b_\varepsilon),\ \forall i, k, t.\}.$$

From the discussion above, there exists a solution to the auxiliary problem. Since θ_{kt} and ε_{kt} are centered in zero with a vanishing probability density for larger values, the solution to the auxiliary problem becomes arbitrarily close to the solution to (2.15) as the bounds $(b_L, b_\theta, b_\varepsilon)$ become larger. If we set $b_L = \bar{b}_L/\beta$, $b_\theta = \beta \bar{b}_\theta$, and $b_\varepsilon = \beta \bar{b}_\varepsilon$ for some fixed $(\bar{b}_L, \bar{b}_\theta, \bar{b}_\varepsilon)$ and $\beta > 0$, the measure of states outside $\mathcal{B}(b_L, b_\theta, b_\varepsilon)$ is vanishing as $\beta \to \infty$, and the solution to the auxiliary problem for $\beta \to \infty$ provides a solution to the dynamic rational inattention problem.

Proof of Proposition 2

Note that the state variables upon which migration decisions depend contain the prior beliefs at the beginning of the period. Indeed, even if we consider one location and two different time periods at which the productivity, population distribution and the agents' preferences are identical but prior beliefs are different, we may still expect agents to acquire different amounts of information, resulting in different beliefs and mobility decisions. The prior belief in turn implicitly depends on the entire history of exogenous states θ^t, endogenous states L^{t-1}, and location decisions l^{t-1}. However, if prior beliefs are "close enough", rationally attentive agents make information acquisition decisions that result in the same posterior beliefs. If one agent is more pessimistic than another about the payoffs in some region, she will be less likely to move there, but conditional on moving, the two agents will have the same posterior beliefs. This property of *locally invariant posteriors* was shown in the context of a static model by Caplin and Dean (2015).[2]

The prior beliefs in some location j at t do depend on the composition of the population inherited from $t-1$. For instance, if the productivity in a neighboring region k was low at $t-1$, the out-flow from k was large and a relatively large number of people in j at t came from k, influencing the shared beliefs in j towards thinking that payoffs in k are low since newcomers from k are pessimistic about k. This dependence of beliefs on the local composition of population is at the core of the model and will deliver central insights. But for property of locally invariant posteriors to hold, we also want to ensure that priors at t do not vary too much with L_{t-1} so that agents will always be able to acquire information that leads to the same posterior beliefs. I will maintain the assumption that it is the case.[3]

[2]Since agents choose the distribution of signals they receive, it is as if they chose their posterior beliefs distributions. Better posterior beliefs increase the expected utility from migration, but have higher entropy. Note that their contribution to agents' utility is separable from the contribution of priors given the information cost function, as long as posterior beliefs can be sustained by priors by Bayes' rule. Therefore if an agent chooses some posterior beliefs, other agents with priors that can also sustain these posteriors will choose the same posteriors.

[3]If Assumption 4 does not hold, then beliefs would depend on all previous population distributions L^{t-1}. I solve the model with 10 regions, allowing for the prior beliefs to vary with (L_{t-1}, L_{t-2}) and posteriors to vary with L_{t-1} and show that for reasonable values of fundamentals and productivity process parameters, the chosen posteriors do not vary with L_{t-1}.

Assumption 4. *The distribution of fundamentals* $(\lambda_j, A_j, B_j, \kappa_{jk})$ *and the productivity process parameters* (ρ, σ_ξ^2) *are such that the ergodic distribution of prior beliefs always lies in the convex hull of optimal posteriors.*

Define $\omega_t = (L_{t-1}, \pi_t, \varepsilon_{it})$. First, following Steiner et al. (2017), note that the problem (2.15) can be written as a control problem with observable states in which the agent must pay a cost for deviating from a default choice rule.

Lemma 2. *A stochastic mobility rule* p *solves the dynamic RI problem if and only if it (together with some default rule* q*) solves:*

$$\max_{q,p} \mathbb{E}\left[\sum_{t=1}^{\infty} \delta^t \left(u_{l_{t-1}l_t t}\left(\theta_t, \omega_t\right) + \lambda_{l_{t-1}}\left(\log q_{l_{t-1}l_t t}\left(\omega_t\right) - \log p_{l_{t-1}l_t t}\left(\theta_t, \omega_t\right)\right)\right)\right]. \tag{B.1.4}$$

where the expectation is with respect to the joint distribution generated by π*,* p*, and* $\gamma(\cdot|\theta)$*.*

Proof. Recall that the dynamic RI problem is equivalent to finding an optimal mobility rule:

$$\max_{p} \mathbb{E}_{\theta^t}\left[\sum_{t=1}^{\infty} \delta^t \left(u_{l_{t-1}l_t}(\theta_t, \omega_t) - I_{l_{t-1}}(\omega_t)\right)\right],$$

where the information cost is expressed as a function of the prior and posterior beliefs $\pi_{jkt} \equiv \pi_{jt|s_t=k}$:

$$I_j\left(\omega_t\right) = \lambda_j\left(H\left(\pi_{jt}\right) - \sum_k q_{jkt}(\omega_t) H\left(\pi_{jkt}\right)\right), \qquad \forall j,$$

and the ex-ante mobility rule is $q_{jkt}(\omega_t) = \int_\theta p_{jkt}(\theta, \omega_t)\pi_{jt}(\theta)d\theta$. As noted by Steiner et al. (2017), by symmetry of the entropy, instead of expressing the information cost as a function of the uncertainty about θ_t before and after receiving the signal, we can rewrite it as a function of the uncertainty about the distribution of mobility choices l_t before and after observing θ_t. Before observing θ_t, l_t is distributed according to p, while once the signal is received, l_t is degenerate, so the entropy upon receiving $s_t = l_t$ is $-\log p_{l_{t-1}l_t t}(\theta_t, \omega_t)$.

We also use the properties of property of properness of the entropy. Properness implies that for any random variable X with finite support S and distribution $g(x) \in \Delta(S)$,

$$H(g) = - \max_{h \in \Delta(S)} \mathbb{E}_g \left[\log h(x) \right]$$

To interpret the properness property, we can consider an agent who believes that X is distributed according to g and is offered to report a distribution $h \in \Delta(S)$ before observing the realization of X, with the promise of a reward of $\log h(x)$ if the realized value is x. Properness implies that the truthful report $h = g$ maximizes the expected reward. The use of properness relies on the information cost being proportional to the reduction in entropy. Applying properness to the distribution p of l_t with finite support J, we get:

$$H(p) = - \max_{q \in \Delta(J)} \mathbb{E}_p \left[\log q(x) \right]$$

This implies that the information cost can be expressed as:

$$\begin{aligned} I_{l_{t-1}}(\omega_t) &= \lambda_{l_{t-1}} \left(H(p) - \mathbb{E}_p \left[-\log p_{l_{t-1} l_t t}(\theta_t, \omega_t) \right] \right) \\ &= \max_{q \in \Delta(J)} \mathbb{E}_p \left[\lambda_{l_{t-1}} \left(-\log q_{l_{t-1} l_t t}(\omega_t) + \log p_{l_{t-1} l_t t}(\theta_t, \omega_t) \right) \right] \end{aligned}$$

Substituting this expression for the information cost into (2.15) gives the result in (B.1.4). □

I now show that (2.17) can be obtained as a solution to the control problem. First, I show that given any default rule q, the dynamic logit rule (2.17) solves problem (B.1.4):

Lemma 3. *Given any default rule q, the control problem with fixed q:*

$$\max_p \mathbb{E} \left[\sum_{t=1}^{\infty} \delta^t \left(u_{l_{t-1} l_t t}(\theta_t, \omega_t) + \lambda_{l_{t-1}} \left(\log q_{l_{t-1} l_t t}(\omega_t) - \log p_{l_{t-1} l_t t}(\theta_t, \omega_t) \right) \right) \right], \tag{B.1.5}$$

has for solution:

$$p_{jkt}(\theta_t, \omega_t) = \frac{q_{jkt}(\omega_t) \exp\left(u_{jkt}(\theta_t, \omega_t) + \delta \bar{V}_{kt+1}(\theta_t, \omega_t) \right)^{1/\lambda_j}}{\sum_l q_{jlt}(\omega_t) \exp\left(u_{jlt}(\theta_t, \omega_t) + \delta \bar{V}_{lt+1}(\theta_t, \omega_t) \right)^{1/\lambda_j}} \tag{B.1.6}$$

and the value function satisfies:

$$\bar{V}_{jt}(\theta_{t-1},\omega_{t-1}) = \mathbb{E}_{\theta_t}\left[\left(\sum_l q_{jlt}(\omega_t)\exp\left(u_{jlt}(\theta_t,\omega_t)+\delta\bar{V}_{lt+1}(\theta_t,\omega_t)\right)^{1/\lambda_j}\right)\right].$$

Proof. For a given default rule $q_{jkt}(\omega_t)$, let the continuation value in the control problem for q:

$$\begin{aligned} & V_{l_{t-1}t}(\theta_t,\omega_t) \\ = & \max_{p_\tau}\sum_{\tau=0}^{\infty}\delta^\tau\mathbb{E}_{\theta_\tau}\left[\sum_k p_{l_{\tau-1}k\tau}(\theta_\tau,\omega_\tau)\Big(u_{l_{\tau-1}k\tau}(\theta_\tau,\omega_\tau)\right. \\ & \left.+\lambda_{l_{\tau-1}}\left(\log q_{l_{\tau-1}k\tau}(\omega_\tau)-\log p_{l_{\tau-1}k\tau}(\theta_\tau,\omega_\tau)\right)\Big)\right] \end{aligned}$$

The value V_{jt} satisfies the recursion:

$$\begin{aligned} & V_{jt}(\theta_t,\omega_t) \\ = & \max_{p_t}\mathbb{E}_{\theta_t}\left[\sum_k p_{jkt}(\theta_t,\omega_t)\big(u_{jkt}(\theta_t,\omega_t)+\lambda_j\left(\log q_{jkt}(\omega_t)-\log p_{jkt}(\theta_t,\omega_t)\right)\right. \\ & \left.+\delta\bar{V}_{kt+1}(\theta_t,\omega_t)\big)\right], \end{aligned} \tag{B.1.7}$$

where $\bar{V}_{kt+1}(\theta_t,\omega_t) = \mathbb{E}\left[V_{kt+1}(\theta_{t+1},\omega_{t+1})|\theta_t,\omega_t\right]$.

To solve the maximization problem in (B.1.8), we can write the first-order condition with respect to $p_{jkt}(\theta_t,\omega_t)$:

$$u_{jkt}(\theta_t,\omega_t)+\lambda_j\left(\log q_{jkt}(\omega_t)-(\log p_{jkt}(\theta_t,\omega_t)+1)\right)+\delta\bar{V}_{kt+1}(\theta_t,\omega_t)=\mu_{jkt}(\theta_t,\omega_t),$$

where $\mu_{jt}(\theta_t,\omega_t)$ is the Lagrange multiplier associated with the constraint $\sum_k p_{jkt}(\theta_t,\omega_t) = 1$. Rearranging the first-order condition gives:

$$p_{jkt}(\theta_t,\omega_t) = \exp\left(\log q_{jkt}(\omega_t)-1+1/\lambda_j\left(u_{jkt}(\theta_t,\omega_t)+\delta\bar{V}_{kt+1}(\theta_t,\omega_t)-\mu_{jkt}(\theta_t,\omega_t)\right)\right).$$

Since $\sum_k p_{jkt}(\theta_t, \omega_t) = 1$, it follows that:

$$\begin{aligned} p_{jkt}(\theta_t, \omega_t) &= \frac{\exp\left(\log q_{jkt}(\omega_t) - 1 + 1/\lambda_j \left(u_{jkt}(\theta_t, \omega_t) + \delta \bar{V}_{kt+1}(\theta_t, \omega_t) - \mu_{jkt}(\theta_t, \omega_t)\right)\right)}{\sum_l \exp\left(\log q_{jlt}(\omega_t) - 1 + 1/\lambda_j \left(u_{jlt}(\theta_t, \omega_t) + \delta \bar{V}_{lt+1}(\theta_t, \omega_t) - \mu_{jkt}(\theta_t, \omega_t)\right)\right)} \\ &= \frac{q_{jkt}(\omega_t) \exp\left(u_{jkt}(\theta_t, \omega_t) + \delta \bar{V}_{kt+1}(\theta_t, \omega_t)\right)^{1/\lambda_j}}{\sum_l q_{jlt}(\omega_t) \exp\left(u_{jlt}(\theta_t, \omega_t) + \delta \bar{V}_{lt+1}(\theta_t, \omega_t)\right)^{1/\lambda_j}}. \end{aligned}$$

Substituting into (B.1.8) gives the recursion:

$$\begin{aligned} & \bar{V}_{jt}(\theta_{t-1}, \omega_{t-1}) \\ = \; & \mathbb{E}_{\theta_t}\left[\sum_k p_{jkt}(\theta_t, \omega_t) \left(\sum_l q_{jlt}(\omega_t) \exp\left(u_{jlt}(\theta_t, \omega_t) + \delta \bar{V}_{lt+1}(\theta_t, \omega_t)\right)^{1/\lambda_j}\right)\right], \\ = \; & \mathbb{E}_{\theta_t}\left[\left(\sum_l q_{jlt}(\omega_t) \exp\left(u_{jlt}(\theta_t, \omega_t) + \delta \bar{V}_{lt+1}(\theta_t, \omega_t)\right)^{1/\lambda_j}\right)\right]. \end{aligned}$$

□

To prove that the logit mobility rule in (B.1.6) is a solution to the dynamic RI equilibrium, it only remains to show that the optimal default rule q from the problem is expressed as:

$$q_{jkt}(\omega_t) = \int_\theta p_{jkt}(\theta, \omega_t) \pi_{jt}(\theta) d\theta.$$

This is implied by properness.

Finally, the last step is to show that in the stochastic steady state, the beliefs π can be summarized as a function of the population distribution. This result relies on the property of locally invariant posteriors, documented by Caplin et al. (2019). This property indicates that agents who face the same payoffs with different prior beliefs about the unobserved components of payoffs will choose an information strategy that will lead them to have the same posterior beliefs upon moving. For a given set of optimal posteriors resulting from the information acquisition strategy, all agents with priors that can sustain these posteriors will choose them. Different priors π_j can sustain the same set of posterior beliefs π_{jk} as long as Bayes' rule is satisfied. So for any θ, it must be that:

$$\pi_j(\theta) = \sum_k q_{jk}\pi_{jk}(\theta).$$

Since q_{jk} can be adjusted under the constraints that $q_{jk} \in [0,1]$ and $\sum_k q_{jk} = 1$, there exists a set of priors beliefs and default rules (π_j, q_{jk}) that satisfy Bayes' rule for the same posterior beliefs. This implies that for priors in this range, agents have the same beliefs conditional on receiving a signal $s = k$, but will choose to receive this signal with different probabilities, leading to different migration shares.

Under the assumption that the parameters When the fluctuations

$$p_{jk}(\theta_t, L_{t-1}, \varepsilon_{it}) = \frac{q_{jk}(L_{t-1}, \varepsilon_{it}) \exp\left(u_{jk}(\theta_t, L_{t-1}, \varepsilon_{it}) + \delta \bar{V}_k(\theta_t, L_{t-1})\right)^{1/\lambda_j}}{\sum_l q_{jl}(L_{t-1}, \varepsilon_{it}) \exp\left(u_{jl}(\theta_t, L_{t-1}, \varepsilon_{it}) + \delta \bar{V}_l(\theta_t, L_{t-1})\right)^{1/\lambda_j}}$$

where $q_{jk}(L_{t-1}, \varepsilon_{it}) = \int_\theta p_{jk}(\theta, L_{t-1}, \varepsilon_{it})\pi_j(\theta|L_{t-1})d\theta$ and we define the expected future value as $\bar{V}_k(\theta_t, L_{t-1}) = \mathbb{E}\left[V_k(\theta_{t+1}, L_t, \varepsilon_{t+1})|\theta_t, L_{t-1}\right]$.

The continuation payoffs solve

$$V_j(\theta_t, L_{t-1}, \varepsilon_{it}) = \lambda_j \log\left(\sum_l q_{jl}(L_{t-1}, \varepsilon_{it}) \exp\left(u_{jl}(\theta_t, L_{t-1}, \varepsilon_{it}) + \delta \bar{V}_l(\theta_t, L_{t-1})\right)^{1/\lambda_j}\right),$$

The posterior beliefs about θ_t held by agents in j if they go to k are determined by Bayes' rule as a function of the prior beliefs about θ_t $\pi_j(\theta_t|L_{t-1})$, the probability of receiving the signal $s = k$ conditional on θ_t, $p_{jk}(\theta_t, L_{t-1}, \varepsilon_{it})$ and the unconditional probability of receiving the signal $s = k$, $q_{jk}(L_{t-1}, \varepsilon_{it})$:

$$\pi_{jk}(\theta_t) = \frac{p_{jk}(\theta_t, L_{t-1}, \varepsilon_{it})\pi_j(\theta_t|L_{t-1})}{q_{jk}(L_{t-1}, \varepsilon_{it})}. \tag{B.1.8}$$

The belief about θ_t held by agents in region k at the end of a period after agents from different origins j have shared their information is expressed as:

$$\log \bar{\pi}_k(\theta_t|L_t) = C_{kt} + \sum_j L_{jkt} \log \pi_{jk}(\theta_t) \tag{B.1.9}$$

The shared belief about θ_t held by agents in region k at the end of a period is then used to form a prior belief about θ_{t+1}, using the exogenous AR(1) transition process for local productivity:

$$\pi_k(\theta_{t+1}|L_t) = \int_\theta \bar{\pi}_k(\theta|L_t)\gamma(\theta_{t+1}|\theta)d\theta, \tag{B.1.10}$$

Population at the end of period t in region k is expressed as a function of the population in every region j at the end of $t-1$ and the mobility probabilities:

$$L_{kt} = \sum_j L_{jt-1}\bar{p}_{jk}(\theta_t, L_{t-1}).. \tag{B.1.11}$$

Proof of Proposition 3

After rewriting agents' problem (2.15) as the control problem in Lemma 2, and plugging in the optimal mobility rule conditional on q obtained in Lemma 3, default choice probabilities are determined as the solutions to the following recursive problem:

$$\max_{\{q_{jl}\}} \int_\theta \lambda_j \log \sum_k q_{jk}(\omega_t) \exp\left(\left(\theta_k + A_k - \alpha \log L_{kt} + B_k - \kappa_{jk} + \delta \bar{V}_k(\theta, L_{t-1})\right)/\nu \right.$$
$$\left. + \varepsilon_{ikt}\right)^{1/\lambda_j} \pi_j(\theta|L_{t-1})d\theta$$

$$s.t. \quad \sum_k q_{jk}(\omega_t) = 1, \quad q_{jk}(\omega_t) \geq 0.$$

Note that $\log \sum_k \exp(v_k/\nu) = \mathbb{E}_e\left[\max_k (v_k + \nu e_k)\right] + C$, where $e_k \sim EV1$ and C is a constant (Small and Rosen, 1981). Applying this result to the problem above, we can rewrite the objective function as:

$$\begin{aligned} & \mathbb{E}_\theta \mathbb{E}_e \left[\max_k \left\{ \frac{A_k - \alpha \log L_{kt} + B_k - \kappa_{jk} + \delta \bar{V}_k(\theta, L_{t-1})}{\nu \lambda_j} + \log q_{jk}(\omega_t) \right.\right. \\ & \left.\left. + \frac{\varepsilon_{ikt}}{\lambda_j} + \frac{\theta_k}{\lambda_j \nu} + e_k \right\}\right] + C \\ = \; & \mathbb{E}_\theta \mathbb{E}_e \left[\max_k \left\{ \frac{\mu_{jk}(L_{t-1}) + A_k - \alpha \log L_{kt} + B_k - \kappa_{jk} + \delta \bar{V}_k(\theta, L_{t-1})}{\nu \lambda_j} + \log q_{jk}(\omega_t) \right.\right. \end{aligned}$$

$$
\begin{aligned}
&+\frac{\varepsilon_{ikt}}{\lambda_j}+\underbrace{\frac{\tilde{\theta}_k}{\lambda_j\nu}+e_k}_{\equiv\zeta_j\tilde{e}_k}\Bigg\}\Bigg]+C\\
=\;&\mathbb{E}_{\tilde{e}}\Bigg[\max_k\Bigg\{\frac{\mu_{jk}(L_{t-1})+A_k-\alpha\log L_{kt}+B_k-\kappa_{jk}+\delta\bar{V}_k(\theta,L_{t-1})}{\nu\lambda_j}+\log q_{jk}(\omega_t)\\
&+\frac{\varepsilon_{ikt}}{\lambda_j}+\zeta_j\tilde{e}_k\Bigg\}\Bigg]+C\\
=\;&\log\sum_k q_{jk}(\omega_t)^{1/\zeta_j}\exp\left(\frac{\mu_{jk}(L_{t-1})+A_k-\alpha\log L_{kt}+B_k-\kappa_{jk}+\delta\bar{V}_k(\theta,L_{t-1})}{\nu\lambda_j\zeta_j}+\frac{\varepsilon_{ikt}}{\lambda_j\zeta_j}\right),
\end{aligned}
$$

where I defined $\zeta_j\tilde{e}_k=\frac{\tilde{\theta}_k}{\lambda_j\nu}+e_k$, and $\tilde{\theta}_k=\theta_k-\mu_{jk}(L_{t-1})$ has mean 0 and variance σ_j^2 under the belief distribution π_j. Note that since $e_k\sim EV1$ and we assumed that $\tilde{\theta}_k$ is distributed according to the conjugate of a type 1 extreme value, it follows from Nadarajah (2008) and Marques et al. (2015) that $\zeta_j\tilde{e}_k\sim EV1$, and the variance of this error is:

$$
\text{Var}\left(\frac{\tilde{\theta}_k}{\lambda_j\nu}+e_k\right)=\frac{\sigma_j^2}{\lambda_j^2\nu^2}+\frac{\pi^2}{6},
$$

where π^2 is here the square of the constant $\pi=3.14159...$ I define the shifter ζ_j to be such that $\text{Var}\left(\tilde{e}_k\right)=\frac{\pi^2}{6}$:

$$
\zeta_j=\left(1+\frac{6\sigma_j^2}{\pi^2\lambda_j^2\nu^2}\right)^{\frac{1}{2}}.
$$

The last line of the previous sequence of equalities was obtained after using the property $\log\sum_k\exp\left(v_k/\nu\right)=\mathbb{E}_e\left[\max_k\left(v_k+\nu e_k\right)\right]+C$, this time with $\tilde{e}_k$. We can therefore rewrite the problem of agents as:

$$
\max_{q_{jl}}\log\sum_k q_{jk}(\omega_t)^{1/\zeta_j}\exp\left(\frac{\mu_{jk}(L_{t-1})+A_k-\alpha\log L_{kt}+B_k-\kappa_{jk}+\delta\bar{V}_l(\theta,L_{t-1})}{\nu\lambda_j\zeta_j}+\frac{\varepsilon_{ikt}}{\lambda_j\zeta_j}\right)
$$

$$
s.t.\quad\sum_k q_{jk}(\omega_t)=1,\quad q_{jk}(\omega_t)\geq 0.
$$

Then, the first order condition with respect to $q_{jl}(\omega_t)$ is:

$$\frac{\partial}{\partial q_{jk}(\omega_t)}\Bigg(\log\sum_k q_{jk}(\omega_t)^{1/\zeta_j}\exp\Big(\frac{\mu_{jk}(L_{t-1})+A_k-\alpha\log L_{kt}+B_k-\kappa_{jk}+\delta\bar{V}_k(\theta,L_{t-1})}{\nu\lambda_j\zeta_j}$$
$$+\frac{\varepsilon_{ikt}}{\lambda_j\zeta_j}\Big)+\varphi\left(1-\sum_k q_{jk}(\omega_t)\right)\Bigg)$$
$$=0,$$

where φ is the Lagrange multiplier. Solving for this first order condition, we obtain a closed-form expression for $q_{jk}(\omega_t)$:

$$q_{jl}(\omega_t)=\frac{\exp\left(\left(\mu_{jk}(L_{t-1})+A_k-\alpha\log L_{kt}+B_k-\kappa_{jk}+\delta\bar{V}_k(\theta,L_{t-1})\right)/\nu+\varepsilon_{ikt}\right)^{\frac{1}{\lambda_j(\zeta_j-1)}}}{\sum_l\exp\left(\left(\mu_{jl}(L_{t-1})+A_l-\alpha\log L_{lt}+B_l-\kappa_{jl}+\delta\bar{V}_l(\theta,L_{t-1})\right)/\nu+\varepsilon_{ilt}\right)^{\frac{1}{\lambda_j(\zeta_j-1)}}}.$$

With an expression for $q_{jk}(\omega_t)$ in hand, we can now derive an expression for the mobility probabilities after information acquisition. Recalling the optimal migration rule:

$$\begin{aligned}
& p_{jk}(\theta_t,\omega_t)\\
=\;& \frac{q_{jk}(\omega_t)\exp\left(\left(\theta_{kt}+A_k-\alpha\log L_{kt}+B_k-\kappa_{jk}+\delta\bar{V}_k(\theta,L_{t-1})\right)/\nu+\varepsilon_{ikt}\right)^{\frac{1}{\lambda_j}}}{\sum_l q_{jkl}(\omega_t)\exp\left(\left(\theta_{lt}+A_l-\alpha\log L_{lt}+B_l-\kappa_{jl}+\delta\bar{V}_l(\theta,L_{t-1})\right)/\nu+\varepsilon_{ilt}\right)^{\frac{1}{\lambda_j}}}\\
=\;& \frac{\exp\left(\frac{\theta_{kt}}{\nu}+\left(A_k-\alpha\log L_{kt}+B_k-\kappa_{jk}+\delta\bar{V}_k(\theta,L_{t-1})\right)\frac{\zeta_j}{\nu(\zeta_j-1)}+\frac{\mu_{jk}(L_{t-1})}{\nu(\zeta_j-1)}+\varepsilon_{ikt}\frac{\zeta_j}{\zeta_j-1}\right)^{\frac{1}{\lambda_j}}}{\sum_l\exp\left(\frac{\theta_{lt}}{\nu}+\left(A_l-\alpha\log L_{lt}+B_l-\kappa_{jl}+\delta\bar{V}_l(\theta,L_{t-1})\right)\frac{\zeta_j}{\nu(\zeta_j-1)}+\frac{\mu_{jl}(L_{t-1})}{\nu(\zeta_j-1)}+\varepsilon_{ilt}\frac{\zeta_j}{\zeta_j-1}\right)^{\frac{1}{\lambda_j}}}
\end{aligned}$$

Recognizing the expression of $p_{jk}(\theta_t,\omega_t)$ as the mobility rule that arises from a multinomial logit decision problem, we can rewrite the problem of agents as:

$$\max_k\tfrac{1}{\lambda_j}\left(\tfrac{\theta_{kt}}{\nu}+\left(A_k-\alpha\log L_{kt}+B_k-\kappa_{jk}+\delta\bar{V}_k(\theta,L_{t-1})\right)\tfrac{\zeta_j}{\nu(\zeta_j-1)}+\tfrac{\mu_{jk}(L_{t-1})}{\nu(\zeta_j-1)}+\varepsilon_{ikt}\tfrac{\zeta_j}{\zeta_j-1}\right)$$
$$+e'_{ikt},$$

where $e'_{ikt}\sim EV1$ and $\mathrm{Var}\left(e'_{ikt}\right)=\pi^2/6$. We can then define $\varrho_j\tilde{e}'_{ikt}=\frac{\zeta_j}{\lambda_j(\zeta_j-1)}\varepsilon_{ikt}+e'_{ikt}$. Since we assume that ε_{ikt} is distributed according to the conjugate of a Type 1 extreme

value distribution, this implies that $\tilde{e}'_{ikt} \sim EV1$, if we define ϱ_j so that $\text{Var}(\tilde{e}'_{ikt}) = \pi^2/6$:

$$\varrho_j = \left(1 + \frac{\zeta_j^2}{\lambda_j^2 (\zeta_j - 1)^2}\right)^{\frac{1}{2}}.$$

Therefore, we can rewrite the problem solved by agents as:

$$\max_k \frac{1}{\varrho_j \lambda_j} \left(\frac{\theta_{kt}}{\nu} + \left(A_k - \alpha \log L_{kt} + B_k - \kappa_{jk} + \delta \bar{V}_k(\theta, L_{t-1})\right) \frac{\zeta_j}{\nu(\zeta_j - 1)} + \frac{\mu_{jk}(L_{t-1})}{\nu(\zeta_j - 1)}\right) + \tilde{e}'_{ikt},$$

to which the solution is:

$$p_{jk}(\theta_t, L_{t-1}) = \frac{\exp\left(\theta_{kt} + \frac{\mu_{jk}(L_{t-1})}{\zeta_j - 1} + \left(A_k - \alpha \log L_{kt} + B_k - \kappa_{jk} + \delta \bar{V}_k(\theta, L_{t-1})\right) \frac{\zeta_j}{\zeta_j - 1}\right)^{\frac{1}{\lambda_j \nu \varrho_j}}}{\sum_l \exp\left(\theta_{lt} + \frac{\mu_{jl}(L_{t-1})}{\zeta_j - 1} + \left(A_l - \alpha \log L_{lt} + B_l - \kappa_{jl} + \delta \bar{V}_l(\theta, L_{t-1})\right) \frac{\zeta_j}{\zeta_j - 1}\right)^{\frac{1}{\lambda_j \nu \varrho_j}}}$$

To arrive at the expression in (2.19), define the following constants:

$$\eta_j = \frac{1}{\zeta_j} = \left(1 + \frac{6\sigma_j^2}{\pi^2 \lambda_j^2 \nu^2}\right)^{-\frac{1}{2}} \qquad \phi_j = \frac{\lambda_j \nu \varrho_j (\zeta_j - 1)}{\zeta_j} = \nu \left(1 + \lambda_j^2 (1 - \eta_j)^2\right)^{1/2}.$$

B.1.2 Alternative Assumptions on the Invertibility of Wages and Population

In this Section, I present an extension of the model to accommodate alternative assumptions on the use of received wages and population as a direct signal on the productivity. In the model, I impose that agents do not exploit the wages they receive, or the observed population distribution, to adjust their beliefs about the local productivity. These restrictions deliver tractability, since beliefs about θ_{kt} do not depend about the realization of θ_{kt}. This Section shows that it is possible to allow for wages and the population distribution to perfectly, or partially, reveal the local value of θ_{kt} or its whole value θ_t, and discusses the implications on the equilibrium mobility patterns.

Complete Revelation of Local Productivity

First, I consider the case in which wages would reveal the most information to agents, while agents do not use the population to update this beliefs. This corresponds to a case in which agents can learn the value of θ_{kt} exactly. As a result, the shared beliefs in k at end of t about θ_{kt} are $\bar{\pi}_{k,k}(\hat{\theta}_{kt}|\theta_{kt}) = \mathbb{1}\{\hat{\theta}_{kt} = \theta_{kt}\}$. The beliefs about the productivity in other locations are still determined by local information sharing (2.13). The prior beliefs about θ_{kt+1} in k at the beginning of the period $t+1$ are then $\pi_{k,k}(\hat{\theta}_{kt+1}|\theta_{kt}) = \phi(\hat{\theta}_{kt+1} - \rho\theta_{kt})$, with a mean $\rho\theta_{kt}$ and variance σ^2_ξ.

This implies that the local prior beliefs will be more precise and unbiased than in the baseline model. The positive bias displayed by agents about their current location in the baseline model is a deterrent to mobility. Agents are optimistic about the productivity in their region, leading them to move out with a lower probability. Therefore, if local beliefs are now unbiased, we should expect agents to be more likely to move out, especially when the local productivity is low. However, the higher precision of local beliefs plays as a counterforce to this first effect. Because the precision of their local prior beliefs is now higher, agents have a lower incentive to invest in information about the local productivity. For this reason, the higher precision of the local prior beliefs relative to other locations may be reduced for posterior beliefs. This tends to limit the responsiveness of migration flows to the current local productivity. Overall, it is likely to depend on the parameter values whether we should expect a higher or lower elasticity of out-migration with respect to local wages.

Depending on which of the two forces described above dominates, for any origin j, these precise local priors $\pi_{j,j}$ may lead to more precise posterior beliefs $\pi_{jk,j}$, regardless of the destination k. If that is the case, it will then improve the precision of the shared beliefs $\bar{\pi}_{k,j}$ about θ_{jt} in all locations k. The two counteracting effects discussed above also arise in this case. Shared beliefs, and hence prior beliefs, will be less biased, reducing the predispositions favoring some regions over others.

Partial Revelation of Local Productivity

I now write an extension of the model in which workers update their beliefs upon observing the wage they obtain in the location they chose in period t. I still maintain the assumption that agents do not invert the law of motion of population (2.12) to learn about θ_t. I assume the existence of an additional unobserved local shock τ_{kt} realized in each location k and period t, after mobility has happened, before production occurs. If this shock has a variance equal to zero, wages reveal local productivity perfectly as in the previous case. If this shock τ_{kt} has a positive variance, even upon observing their wage, agents cannot perfectly infer the realization of the local productivity θ_{kt}.

Each period, agents who move from location j to k receive the flow of payoff u_{ijkt}, not inclusive of information costs, given by (2.1). The structure of production is similar to the baseline model, with a unique freely traded homogeneous good chosen as numeraire. The production function takes a Cobb-Douglas form using labor as the single input, so that output in region k at t is given by:

$$y_{kt} = \exp(A_k + \theta_{kt} + \tau_{kt}) L_{kt}^{1-\alpha}, \qquad \alpha > 0, \tag{B.1.12}$$

where θ_{kt} follows the AR(1) process described in (2.3), and τ_{kt} are iid across locations and time, and drawn from a normal distribution $\mathcal{N}\left(0, \sigma_\tau^2\right)$. I assume that τ_{kt} is unobserved. However, contrary to θ_{kt}, agents have no incentive to acquire information about τ_{kt}. This is first because τ_{kt} is realized only after mobility has occurred, so that no information about it can be collected to adjust mobility decisions at t. Second, τ_{kt} is iid over time, so any information about τ_{kt} has no predictive power about its future value.

The first order condition arising from profit maximization yields the following expression for the wage in region k at t:

$$\log w_{kt} = A_k + \theta_{kt} + \tau_{kt} - \alpha \log L_{kt} + \log\left(1 - \alpha\right). \tag{B.1.13}$$

As a result, indirect utility, net of information costs introduced below, can be expressed as

$$u_{ijkt} = A_k + \theta_{kt} + \tau_{kt} - \alpha \log L_{kt} + B_k - \kappa_{jk} + \nu\varepsilon_{ikt}. \tag{B.1.14}$$

Given that local productivity shocks are the only variables that agents are trying to observe, agents still hold beliefs about the cross-sectional distribution of productivity:

$$\pi_{it}(\theta) = \Pr(\theta_t = \theta | i, t), \qquad \forall \theta \in \mathbb{R}^J. \tag{B.1.15}$$

Figure B.1.1 represents the timeline of events. In addition to the baseline model, the local shock τ_{kt} is realized once agents have made location decisions. Production occurs according to (B.1.12). The wage is then observed, and agents in k update their beliefs about θ_{kt} according to Bayes' rule. Note the observation of w_{kt} can only be used by agents to update their beliefs about productivity in the current location k, since productivity shocks are independent across locations. Denoting by $\pi_{j,k}(\theta_{kt})$ the marginal belief about θ_{kt} held by agents in k after information sharing, the updating for any possible $\theta_{kt} = \hat{\theta}_{kt}$ upon observing a realization of wages $w_{kt} = \hat{w}_{kt}$ is given by

$$\Pr(\theta_{kt} = \hat{\theta}_{kt} | w_{kt} = \hat{w}_{kt}) = \frac{\Pr(w_{kt} = \hat{w}_{kt} | \theta_{kt} = \hat{\theta}_{kt}) \Pr(\theta_{kt} = \hat{\theta}_{kt})}{\Pr(w_{kt} = \hat{w}_{kt})}.$$

Assuming that $\pi_{j,k}(\theta_{kt})$ is the pdf of a normal distribution with some mean $\mu_{j,k}$ and variance $\sigma^2_{j,k}$, the resulting belief is expressed as

$$\theta_{kt} | w_{kt} = \hat{w}_{kt} \sim \mathcal{N}\left(\frac{\sigma^2_{j,k} \left(\hat{w}_{kt} - A_k + \alpha \log L_{kt} - \log(1-\alpha) \right) + \sigma^2_\tau \mu_{j,k}}{\sigma^2_{j,k} + \sigma^2_\tau}, \left(\sigma^{-2}_{j,k} + \sigma^{-2}_\tau \right)^{-1} \right). \tag{B.1.16}$$

In contrast to the baseline model, the end-of-period posterior beliefs about θ_{kt} depend on the actual realization of w_{kt}, which itself depends on the actual realization of (θ_{kt}, τ_{kt}). We can rewrite the final posterior beliefs, after observation of local wages, as

$$\pi_j(\theta_{kt} | L_{t-1}, \theta_{jt-1}, \tau_{jt-1})$$

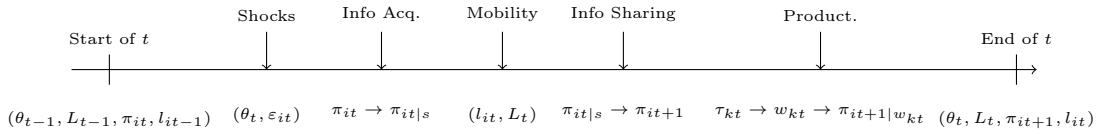


Figure B.1.1: Sequence of actions during any period t

The information acquisition problem is unchanged, and agents pay the information costs (2.9). The updating of beliefs after having the signal $s = k$ is given by

$$\pi_{jk}(\theta_t) = \frac{\bar{p}_{jk}(\theta_t, L_{t-1}, \theta_{jt-1}, \tau_{jt-1})\pi_j(\theta_t|L_{t-1}, \theta_{jt-1}, \tau_{jt-1})}{q_{jk}(L_{t-1}, \theta_{jt-1}, \tau_{jt-1})}. \tag{B.1.17}$$

The population in any location k after mobility has occurred is then $L_{kt} = \sum_j L_{jkt}$, where

$$L_{jkt} = L_{jt-1}\bar{p}_{jk}(\theta_t, L_{t-1}, \theta_{jt-1}, \tau_{jt-1}), \tag{B.1.18}$$

The belief about θ_t held by agents in region k at the end of a period after agents from different origins j have shared their information is expressed exactly as in the baseline,

$$\log \bar{\pi}_k(\theta_t|L_t) = C_{kt} + \sum_j L_{jkt} \log \pi_{jk}(\theta_t). \tag{B.1.19}$$

We can then write the analogous to Proposition 2, expressing the mobility rule for each preference shock ε_{it}, now as a function of the previous local productivity shocks θ_{jt-1}, and τ_{jt-1}.

Proposition 6. *In the stochastic steady state, for each agent located in j at $t-1$, the optimal mobility rule $p_{jk}(\theta_t, L_{t-1}, \theta_{jt-1}, \tau_{jt-1}, \varepsilon_{it})$ can be expressed as*

$$\begin{aligned} & p_{jk}(\theta_t, L_{t-1}, \theta_{jt-1}, \tau_{jt-1}, \varepsilon_{it}) \\ = \; & \frac{q_{jk}(L_{t-1}, \theta_{jt-1}, \tau_{jt-1}, \varepsilon_{it}) \exp\left(u_{jk}(\theta_t, L_{t-1}, \varepsilon_{it}) + \delta \bar{V}_k(\theta_t, L_{t-1})\right)^{1/\lambda_j}}{\sum_l q_{jl}(L_{t-1}, \theta_{jt-1}, \tau_{jt-1}, \varepsilon_{it}) \exp\left(u_{jl}(\theta_t, L_{t-1}, \varepsilon_{it}) + \delta \bar{V}_l(\theta_t, L_{t-1})\right)^{1/\lambda_j}} \end{aligned} \tag{B.1.20}$$

where $q_{jk}(L_{t-1}, \theta_{jt-1}, \tau_{jt-1}, \varepsilon_{it}) = \int_\theta p_{jk}(\theta, \theta_{jt-1}, \tau_{jt-1}, L_{t-1}, \varepsilon_{it})\pi_j(\theta|L_{t-1}, \theta_{jt-1}, \tau_{jt-1})d\theta$ *and we define the expected future value as*

$\bar{V}_k(\theta_t, L_{t-1}, \theta_{jt-1}, \tau_{jt-1}) = \mathbb{E}\left[V_k(\theta_{t+1}, L_t, \varepsilon_{it+1}, \theta_{jt}, \tau_{jt})|\theta_t, L_{t-1}, \theta_{jt-1}, \tau_{jt-1}\right]$. *The continuation payoffs solve*

$$V_j(\theta_t, L_{t-1}, \theta_{jt-1}, \tau_{jt-1}, \varepsilon_{it}) = \tag{B.1.21}$$
$$\lambda_j \log\left(\sum_l q_{jl}(L_{t-1}, \theta_{jt-1}, \tau_{jt-1}, \varepsilon_{it}) \exp\left(u_{jl}(\theta_t, L_{t-1}, \varepsilon_{it}) + \delta \bar{V}_l(\theta_t, L_{t-1}, \theta_{jt-1}, \tau_{jt-1})\right)^{1/\lambda_j}\right),$$

and population and beliefs follow the laws of motion (2.12) - (2.14).

Proof. See Appendix B.1.1. □

After imposing Assumptions 1 and 2 on the equilibrium distribution of beliefs and preference shocks, we can write the equilibrium aggregate mobility rule in a similar fashion to Proposition 3.

Proposition 7. *Under Assumptions 1 and 2, the average mobility rule in the presence of information frictions and preference heterogeneity is given by*

$$\begin{aligned} & \bar{p}_{jk}(\theta_t, L_{t-1}, \theta_{jt-1}, \tau_{jt-1}) \\ = \; & \frac{\exp\left(\eta_j \chi_{jk}(\theta_t, L_{t-1}, \theta_{jt-1}, \tau_{jt-1}) + \bar{v}_{jk}(\theta_t, L_{t-1}, \theta_{jt-1}, \tau_{jt-1})\right)^{\frac{1}{\phi_j}}}{\sum_l \exp\left(\eta_j \chi_{jl}(\theta_t, L_{t-1}, \theta_{jt-1}, \tau_{jt-1}) + \bar{v}_{jl}(\theta_t, L_{t-1}, \theta_{jt-1}, \tau_{jt-1})\right)^{\frac{1}{\phi_j}}}, \end{aligned} \tag{B.1.22}$$

where $\chi_{jk}(\theta_t, L_{t-1}, \theta_{jt-1}, \tau_{jt-1}) = \mu_{jk}(L_{t-1}, \theta_{jt-1}, \tau_{jt-1}) - \theta_{kt}$ *is the expectational error made by agents in* j *about* θ_{kt}*, while the continuation payoffs solve*

$$\begin{aligned} & V_j(\theta_t, L_{t-1}, \theta_{jt-1}, \tau_{jt-1}) \\ = \; & \phi_j \log\left(\sum_l \exp\left(\eta_j \chi_{jl}(\theta_t, L_{t-1}, \theta_{jt-1}, \tau_{jt-1}) + \bar{v}_{jl}(\theta_t, L_{t-1}, \theta_{jt-1}, \tau_{jt-1})\right)^{\frac{1}{\phi_j}}\right), \end{aligned} \tag{B.1.23}$$

with $\phi_j = \nu\left(1 + \lambda_j^2 (1 - \eta_j)^2\right)^{1/2}$ *and* $\eta_j = \left(1 + \frac{6\sigma_j^2}{\pi^2 \lambda_j^2 \nu^2}\right)^{-1/2} \in (0, 1)$*, and* π *is the constant* $\pi = 3.1415...$

When the information cost tends to zero, $\mu_{jk}(L_{t-1}) \to \theta_{kt}$ and $\phi_j \to \nu$ so that the model reduces to a preference-based migration model:

$$\bar{p}_{jk}(\theta_t, L_{t-1}) = \frac{\exp\left(\bar{v}_{jk}(\theta_t, L_{t-1})\right)^{1/\nu}}{\sum_l \exp\left(\bar{v}_{jl}(\theta_t, L_{t-1})\right)^{1/\nu}}. \tag{B.1.24}$$

When the dispersion of preferences ν tends to zero, the solution becomes

$$\begin{aligned} & \bar{p}_{jk}(\theta_t, L_{t-1}, \theta_{jt-1}, \tau_{jt-1}) \\ = \; & \frac{\exp\left(\rho_j \chi_{jk}(\theta_t, L_{t-1}, \theta_{jt-1}, \tau_{jt-1}) + \bar{v}_{jk}(\theta_t, L_{t-1}, \theta_{jt-1}, \tau_{jt-1})\right)^{\frac{1}{\psi_j}}}{\sum_l \exp\left(\rho_j \chi_{jl}(\theta_t, L_{t-1}, \theta_{jt-1}, \tau_{jt-1}) + \bar{v}_{jl}(\theta_t, L_{t-1}, \theta_{jt-1}, \tau_{jt-1})\right)^{\frac{1}{\psi_j}}}, \end{aligned} \tag{B.1.25}$$

where $\psi_j = \lambda_j (1 - \rho_j)$ *and* $\rho_j = \left(1 + \frac{6\sigma_j^2}{\pi^2 \lambda_j^2}\right)^{-1/2} \in (0, 1)$.

The implications of letting workers update their beliefs after observing wages depends on the the parameters of the AR(1) process and the variance of the idiosyncratic shocks τ_{kt}. If the persistence ρ is small, then any information about θ_{kt} obtained after migration at t has little value for future mobility choices. At the limit where $\rho = 0$, updating beliefs after observing wages or not does not affect the equilibrium behavior. If $\rho > 0$ but σ_τ^2 is large in comparison to σ_ξ^2, then wages w_{kt} contain little information about θ_{kt}. As shown by (B.1.16), as the variance σ_{jk}^2 of agents' beliefs about θ_{kt} becomes small relative to σ_τ^2, the beliefs are almost unchanged after observing the wages. For any intermediate value of ρ and σ_τ^2, however, observing wages will make local beliefs more precise and influence the equilibrium behavior.

Identifying the variance of the shock τ_{kt}, σ_τ^2, along with the persistence, ρ, and variance, σ_ξ^2, of the innovation of the AR(1) process, is possible by computing the variance and autocovariances of the adjusted wages,

$$\tilde{w}_{kt} = \log w_{kt} + \alpha \log L_{kt} - \log(1 - \alpha) = A_k + \theta_{kt} + \tau_{kt}.$$

It is easy to show that the parameters are identified by the following expressions,

$$\frac{\text{Cov}\left(\tilde{w}_{kt}, \tilde{w}_{kt-2}\right)}{\text{Cov}\left(\tilde{w}_{kt}, \tilde{w}_{kt-1}\right)} = \rho,$$

$$\frac{1-\rho^2}{\rho}\text{Cov}\left(\tilde{w}_{kt}, \tilde{w}_{kt-1}\right) = \sigma_\xi^2,$$

$$\text{Var}\left(\tilde{w}_{kt}\right) - \frac{\sigma_\xi^2}{1-\rho^2} = \sigma_\tau^2.$$

Complete Revelation of the Productivity Vector

Now turning to the case that allows workers to exploit the complete structure of the model to infer information about the productivity vector θ_t. As I discuss in Section 2.2.4, agents do use the mobility rule $\bar{p}_{jk}(\theta_t, L_{t-1})$ to compute the wages in every location. After mobility has occurred, agents observe the population distribution L_t. Recall that the law of motion of population is given by

$$L_{jkt} = L_{jt-1}\bar{p}_{jk}(\theta_t, L_{t-1}),$$

so that upon observing L_{jkt} for all locations j, k, and with the knowledge of L_{t-1} and the mobility rule $\bar{p}$, agents can in principle invert this equation to recover θ_t. Their ability to exactly infer θ_t depends on whether the function $\theta_t \mapsto \bar{p}(\theta_t, L_{t-1})$ associating to each productivity vector, a vector of moving probabilities, conditional on a population distribution, is injective. If it is not, this inversion is complex. For the sake of this argument, let's consider the case in which it is, so that θ_t is completely revealed to all agents at the end of period t.

This implies that the prior beliefs $\pi_j(\theta_t)$, in period t in every location j, are identical and reflect the AR(1) process governing the evolution of θ_t,

$$\theta_t \sim \mathcal{N}\left(\rho\theta_{t-1}, \sigma_\xi^2\right).$$

This makes social learning nullified in this context, since all information is revealed to all agents. Despite having the same prior beliefs, agents however decide to acquire different information depending on their location and on their idiosyncratic shocks. There is therefore

still a meaningful role for information frictions in this context. In particular, agents will acquire less information about distant locations, leading to the heterogeneous migration elasticity results described in Section 2.5.1.

B.1.3 Solution Algorithm

We assume that the stochastic productivity component in each region follows an AR(1) process:

$$\theta_{kt} = \rho\theta_{kt-1} + \varepsilon_{kt}, \qquad \varepsilon_{kt} \sim \mathcal{N}(0, \sigma_\varepsilon),$$

with all ε_{kt} are iid across time periods and regions. With the current specification, the ergodic distribution of θ_{kt} is a normal distribution with mean 0 and variance $\sigma_\varepsilon/(1-\rho^2)$. We make the approximation that p_{jk}, π_{jk} and V_j take the following forms:

$$p_{jk}(\theta'^i, L^i) = g_{jk,L}(L^i)\prod_l g_{jk,l}(\theta'^i_l)$$

$$\pi_{jk}(\theta'^i) = \prod_l \pi_{jk,l}(\theta'^i_l)$$

$$V_j(\theta'^i, L^i) = V_{j,L}(L^i)\prod_l V_{j,l}(\theta'^i_l)$$

Moreover, we assume that the l-partial beliefs $\pi_{jk,l}(\theta'^i_l)$ are pdfs of normal distribution:

$$\pi_{jk,l}(\theta'^i_l) = \frac{1}{\sqrt{2\pi\sigma^2_{jk,l}}}\exp\left(-\frac{\left(\theta^i_l - \mu_{jk,l}\right)^2}{2\sigma^2_{jk,l}}\right)$$

and define:

$$V_{k,L}(L'^i) = \sum_{m,j}\beta^{V,L}_{k,mj}\log L^i_{mj}, \qquad V_{k,l}(\theta'^i_l) = \beta^V_{1k,l}\theta'^i_l,$$

$$g_{jk,L}(L^i) = \exp\left(\sum_{m,j}\beta^{g,L}_{jk,mj}\log L^i_{mj}\right), \qquad g_{jk,l}(\theta'^i_l) = \exp\left(\beta^g_{1jk,l}\theta'^i_l + \beta^g_{2jk,l}\left(\theta'^i_l\right)^2\right).$$

The sample of states consists of N_l population vectors (each of length J^2), and N_s productivity vectors (each of length J). The number of possible values for θ_k is set to

$n_s = 2$. There are therefore $N = N_s N_l$ states in total, and we denote by i the index of one such state.

1. Guess partial value functions $V^{(0)}_{j,L}$, $V^{(0)}_{j,l}$, a partial choice rule $g^{(0)}_{jk,L}$, $g^{(0)}_{jk,l}$, and partial beliefs $\pi^{(0)}_{jk,l}$. Use the result of last loop if there is one. Compute the next period population that would arise from any pair (θ'^i, L^i):

$$L'^i_{jk} = p_{jk}(\theta'^i, L^i) L^i_j.$$

Define the gross payoff, exponential total payoff, and shifted exponential payoff of action $j \to k$:

$$u_{jk}(\theta'^i, L^i) = A_k + \theta'^i_k - \alpha \log L'^i_k + B_k - \kappa_{jk}$$

2. Compute the end-of period homogenized beliefs for people in j that the current state is θ'^i if the end-of-period population is L^i:

$$\bar{\pi}_j(\theta'^i | L^i) = \prod_{l=1}^{J} \bar{\pi}_{j,l}(\theta'^i_l | L^i),$$

and the l-partial homogenized belief takes the form:

$$\begin{aligned} \bar{\pi}_{j,l}(\theta'^i_l | L^i) &= \frac{1}{C_{j,l}(L^i)} \prod_{m=1}^{J} \pi_{mj,l}(\theta'^i_l)^{\frac{L^i_{mj}}{L^i_j}} \\ &= \frac{1}{\sqrt{2\pi \bar{\sigma}_{j,l}(L^i)^2}} \exp\left(- \frac{\left(\theta'^i_l - \bar{\mu}_{j,l}(L^i)\right)^2}{2\bar{\sigma}_{j,l}(L^i)^2} \right) \end{aligned}$$

with

$$\bar{\sigma}_{j,l}(L^i)^2 = \left(\frac{1}{L^i_j} \sum_{m=1}^{J} \frac{L^i_{mj}}{\sigma^2_{mj,l}} \right)^{-1}, \qquad \bar{\mu}_{j,l}(L^i) = \bar{\sigma}_{j,l}(L^i)^2 \frac{1}{L^i_j} \sum_{m=1}^{J} \frac{L^i_{mj}}{\sigma^2_{mj,l}} \mu_{mj,l}.$$

Then compute the belief for people in j that the *next* period state is θ'^i if the end-of-period population is L^i:

$$E\bar{\pi}_j(\theta'^i|L^i) = \prod_{l=1}^{J} E\bar{\pi}_{j,l}(\theta_l'^i|L^i),$$

and the l-partial homogenized belief takes the form:

$$\begin{aligned} E\bar{\pi}_{j,l}(\theta_l'^i|L^i) &= \int_{-\infty}^{\infty} \phi_\varepsilon\left(\theta_l'^i - \rho\theta_l\right) \bar{\pi}_{j,l}(\theta_l|L^i)d\theta_l \\ &= \frac{1}{\sqrt{2\pi\hat{\sigma}_{j,l}(L^i)^2}} \exp\left(-\frac{\left(\theta_l'^i - \hat{\mu}_{j,l}(L^i)\right)^2}{2\hat{\sigma}_{j,l}(L^i)^2}\right) \end{aligned}$$

with

$$\hat{\sigma}_{j,l}(L^i)^2 = \rho^2\bar{\sigma}_{j,l}(L^i)^2 + \sigma_\varepsilon^2, \qquad \hat{\mu}_{j,l}(L^i) = \rho\bar{\mu}_{j,l}(L^i)$$

3. Solve for the unconditional moving probability $q_{jk}(L^i)$:

$$\begin{aligned} & q_{jk}(L^i) \\ = \; & \mathbb{E}_{\theta'}\left[p_{jk}(\theta', L^i)E\bar{\pi}_j(\theta'|L^i)\right] \\ = \; & g_{jk,L}(L^i)\prod_{l=1}^{J}\int_{-\infty}^{\infty} g_{jk,l}(\theta_l)E\bar{\pi}_{j,l}(\theta_l|L^i)d\theta_l \\ = \; & C_{qj}(L^i)g_{jk,L}(L^i)\exp\left(\sum_l \frac{\beta_{1jk,l}^g\left(\hat{\mu}_{jl}(L^i) + \frac{1}{2}\hat{\sigma}_{jl}(L^i)^2\beta_{1jk,l}^g\right) + \beta_{2jk,l}^g\hat{\mu}_{jl}(L^i)^2}{1 - 2\hat{\sigma}_{j,l}(L^i)^2\beta_{2jk,l}^g}\right) \end{aligned}$$

where $C_{qj}(L^i)$ is a normalizing constant to ensure that $\sum_k q_{jk}(L^i) = 1$.

4. Update the value function by iterating on the Bellman equation. First, compute the expected value in the next period if the current productivity and previous population is (θ'^i, L^i):

$$\begin{aligned} EV_k(\theta'^i, L^i) &= \mathbb{E}_{\theta''}\left[V_k(\theta'', L'^i)|\theta'^i\right] \\ &= V_{k,L}(L'^i) + \sum_l \beta_{1k,l}^V \int_{-\infty}^{\infty} \theta_l\phi_\varepsilon(\theta_l - \rho\theta_l'^i)d\theta_l \end{aligned}$$

$$= \quad V_{k,L}(L^{'i}) + \rho \sum_l \beta^V_{1k,l} \theta^{'i}_l$$

Define the convenient transformed payoff $z_{jk}(\theta^{'i}, L^i)$:

$$z_{jk}(\theta^{'i}, L^i) = \exp\left(u_{jk}(\theta^{'i}, L^i) + \delta EV_k(\theta^{'i}, L^i)\right)^{1/\lambda_j}$$

Then use the Bellman equation:

$$V_j(\theta^{'i}, L^i) = \bar{\gamma} + \lambda_j \log\left(\sum_k q_{jk}(L^i) z_{jk}(\theta^{'i}, L^i)\right).$$

To recover the partial values, run the following regression for each k:

$$V_k(\theta^{'i}, L^i) = \sum_l \left(\beta^V_{1k,l} \theta^{'i}_l\right) + \sum_{m,j} \beta^{V,L}_{k,mj} \log L^i_{mj} + \varepsilon^i_k,$$

and update:

$$V_{k,L}(L^{'i}) = \sum_{m,j} \beta^{V,L}_{k,mj} \log L^i_{mj}, \qquad V_{k,l}(\theta^{'i}_l) = \beta^V_{1k,l} \theta^{'i}_l.$$

Return to the definition of $EV_k(\theta^{'i}, L^i)$ with these updated $V_{k,L}$ and $V_{k,l}$ and keep looping until $V_k(\theta^{'i}, L^i)$ has converged.

5. Update the decision rule $p_{jk}(\theta^{'i}, L^i)$:

$$p_{jk}(\theta^{'i}, L^i) = \frac{q_{jk}(L^i) z_{jk}(\theta^{'i}, L^i)}{\sum_l q_{jl}(L^i) z_{jl}(\theta^{'i}, L^i)}$$

To recover the partial choice rule, run the following regression for each jk:

$$\log p_{jk}(\theta^{'i}, L^i) = \sum_l \left(\beta^g_{1jk,l} \theta^{'i}_l + \beta^g_{2jk,l} \left(\theta^{'i}_l\right)^2\right) + \sum_{m,j} \beta^{g,L}_{jk,mj} \log L^i_{mj} + \varepsilon^i_{jk},$$

and update:

$$g_{jk,L}(L^i) = \exp\left(\sum_{m,j} \beta^{g,L}_{jk,mj} \log L^i_{mj}\right), \qquad g_{jk,l}(\theta_l'^i) = \exp\left(\beta^g_{1jk,l}\theta_l'^i + \beta^g_{2jk,l}\left(\theta_l'^i\right)^2\right).$$

6. Update the partial beliefs

$$\begin{aligned}\pi_{jk,l}(\theta_l'^i) &= g_{jk,l}(\theta_l'^i)\prod_{i=1}^{N_l} E\bar{\pi}_{j,l}(\theta_l'^i|L^i)^{\frac{1}{N_l}} \\ &= \frac{1}{\sqrt{2\pi\sigma_{jk,l}'^2}}\exp\left(-\frac{\left(\theta_l'^i - \mu'_{jk,l}\right)^2}{2\sigma_{jk,l}'^2}\right)\end{aligned}$$

with the updated mean and variances

$$\sigma_{jk,l}'^2 = \left(\frac{1}{N_l}\sum_{i=1}^{N_l}\frac{1}{\hat{\sigma}_{jl}(L^i)^2} - 2\beta^g_{2jk,l}\right)^{-1} \qquad \mu'_{jk,l} = \sigma_{jk,l}'^2\left(\beta_{1jk,l} + \frac{1}{N_l}\sum_{i=1}^{N_l}\frac{\hat{\mu}_{j,l}(L^i)}{\hat{\sigma}_{jl}(L^i)^2}\right)$$

7. Compare the updated values of $p_{jk}(\theta'^i, L^i)$ and $\pi_{jk}(\theta'^i)$ to their previous ones, and if they are not very similar, go back to step 1 with the new values of $V_{j,L}$, $V_{j,l}$, partial choice rule $g_{jk,L}$, $g_{jk,l}$, and partial beliefs $\pi_{jk,l}$.

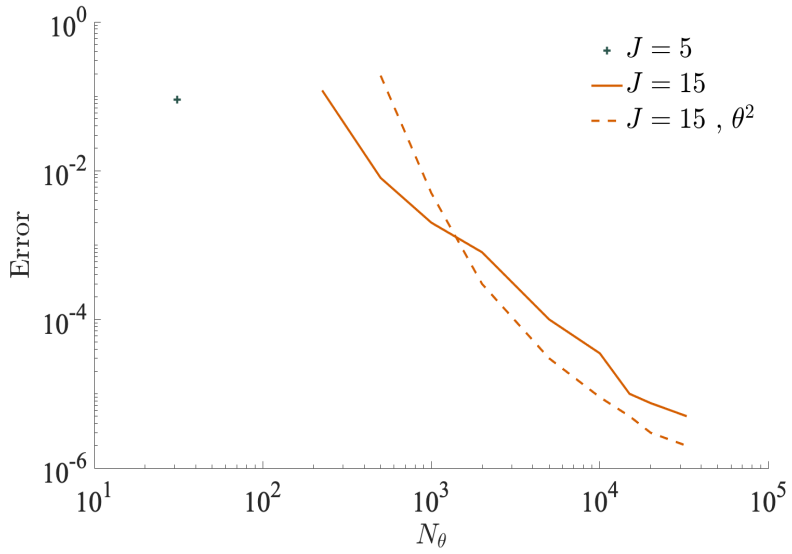
B.1.4 Accuracy of the Solution Algorithm

Binary productivity process

For few regions and a simple productivity process, it is possible to compute an "exact" solution of the model.

Figure B.1.2 shows that the simulation error decreases fast with the number N_θ of states (θ, L) drawn for the solution of the model.

Figure B.1.2: Simulation Error



To compute the exact solution of the model , assume that instead of following an AR(1) process, the productivity is discrete and binary, so that $\theta_{jt} \in \{0, 1\}$, and restrict the number of regions to be at most 15, $J < 15$. Draw $N_L = 1000$ population samples. I then simulate the "almost exact" model x^*, where "almost exact" indicates that the solution is still obtained with a sample of population values. Then compare to the approximation $\hat{x}$, and compute the error as the mean of $|(x^* - \hat{x})/x^*|$. Figure B.1.2 indicates that the error is lower than 0.1% as soon as $N_\theta = 1000$. The comparison of the solid and dashed line illustrates that allowing for higher order terms in the approximation of the value function have little effect on the overall precision.

Discretized AR(1) Process

In this section, I discuss the consequence of approximating beliefs about the vector of productivity θ_t by independent distributions of θ_{kt}, $k = 1, \ldots, J$. I consider the case of normally distributed beliefs

To compare the approximated solution to the exact one, I restrict the analysis to a small number of regions, namely $J = 4$, and assume that preferences are homogeneous across agents, so that $\nu = 0$. I discretize the AR(1) process into $n_\theta = 20$ possible values according to

Tauchen's method. This implies that the total number of possible productivity combinations is $N_\theta = n_\theta^J = 160,000$. I then draw $N_L = 1000$ population samples $L^i = (L^i_{jk})_{j,k=1,\dots J}$ for $i = 1, \dots, N_L$, such that $L_{jk} \geq 0$ and $\sum_{j,k} L^i_{jk} = \bar{L}$.

The algorithm for finding the exact solution of the model in this configuration is as follows.

1. Guess an expression $p^{(0)}_{jk}$ for the moving probability between every pair of locations j, k, and for each state $(\theta^{i'}, L^i)$. One simple guess is:

$$p^{(0)}_{jk}(\theta^{i'}, L^i) = \frac{\exp(u_{jk}(\theta^{i'}, L^i))^{1/\lambda_j}}{\sum_l \exp(u_{jl}(\theta^{i'}, L^i))^{1/\lambda_j}},$$

   where the flow of utility is determined in general by

$$u_{jk}(\theta^{i'}, L^i) = \theta^{i'}_k - \alpha \log L'^i_k + A_k + B_k - \kappa_{jk},$$

   and the end-of-period population is given by $L'^i_k = \sum_j L^i_{jk} p_{jk}(\theta^{i'}, L^i)$. For the first guess, I set $L'^i = L^i$, since no expression of p_{jk} is available yet.

2. Guess an expression $\pi^{(0)}_j$ for the prior beliefs in each region about the productivity vector, conditional on the population vector. I set the belief about $\theta^{i'}$ to be the product of the marginal probabilities

$$\pi^{(0)}_j(\theta^{i'}|L^i) = \prod_k P^*(\theta^{i'}_k),$$

   where P^* is the steady state distribution of the discretized AR(1) process, satisfying $P^*\Gamma = P^*$, with $\Gamma(\cdot)$ the transition matrix of the Markov Chain representing the discretized AR(1) process.

3. Using the current guesses of moving probabilities $p^{(n)}_{jk}$ and beliefs $\pi^{(n)}_j$, construct the ex-ante moving probabilities

$$q^{(n)}_{jk}(L^i) = \sum_{i'} p^{(n)}_{jk}(\theta^{i'}, L^i) \sum_{i''} \pi^{(n)}_j(\theta^{i''}|L^i)\Gamma(\theta^{i'}|\theta^{i''})$$

4. Solve for the value function $V_j^{(n)}(\theta^i, L^{i'})$ as a fixed point of the Bellman equation

$$V_j^{(n)}(\theta^{i'}, L^i) = \lambda_j \log \sum_k q_{jk}^{(n)}(L^i) \exp\left(u_{jk}(\theta^{i'}, L^i) + \delta \bar{V}_k(\theta^{i'}, L^i)\right)^{1/\lambda_j}$$

where the expected value is

$$\bar{V}_k(\theta^{i'}, L^i) = \sum_{i''} V_k(\theta^{i''}, \tilde{L}'^i)\Gamma(\theta^{i''}|\theta^{i'}),$$

where $\tilde{L}'^i$ is the closest to L'^i within the sample of population vectors.

5. Update the moving probability

$$p_{jk}^{(n+1)}(\theta^{i'}, L^i) = \frac{q_{jk}^{(n)}(L^i) \exp(u_{jk}(\theta^{i'}, L^i) + \delta \bar{V}_k^{(n)}(\theta^{i'}, L^i))^{1/\lambda_j}}{\sum_l q_{jl}^{(n)}(L^i) \exp(u_{jl}(\theta^{i'}, L^i) + \delta \bar{V}_l^{(n)}(\theta^{i'}, L^i))^{1/\lambda_j}}.$$

6. Update posterior beliefs

$$\pi_{jk}^{(n+1)}(\theta^{i'}|L^i) = \frac{p_{jk}^{(n+1)}(\theta^{i'}, L^i) \sum_{i''} \pi_j^{(n)}(\theta^{i''}|L^i)\Gamma(\theta^{i'}|\theta^{i''})}{q_{jk}^{(n)}(L^i)}$$

7. Update shared beliefs

$$\log \bar{\pi}_k^{(n+1)}(\theta^{i'}|L^i) = C_k(L^i) + \sum_j L'^i_{jk} \log \pi_{jk}^{(n+1)}(\theta^{i'}|L^i).$$

B.2 Appendix to Section 2.3

B.2.1 Construction of the Main Sample

For every formal job and year, I exploit information on the day of accession into the job and the day of separation (if either of them took place during the corresponding year), the average monthly wage, the number of hours stipulated in the contract, the 2-digit occupation (according to the *Classificação Brasileira de Ocupações*, CBO), and certain characteristics of both the plant at which the worker is employed and of the worker herself. Specifically, I use information on the micro and mesoregion in which the plant is located and its main 2-digit industry of production (according to the *Classificação Nacional de Atividades Econômicas*, CNAE), as well as information on the workers' gender, age, and level of education.[4]

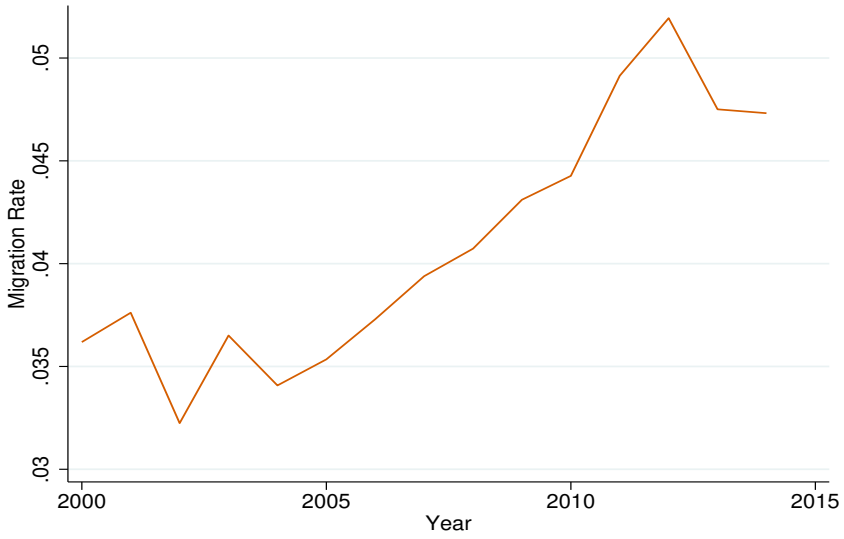
Because RAIS only contains information on the formal employment of workers in Brazil, I have no information on the location of workers that do not hold a formal job in a given year. These workers may be employed in the informal sector, self-employed, unemployed, or out of the labor force. Given that my results will naturally capture only the incidence of informational frictions for migration decisions of workers employed in the formal sector, I limit the analysis to workers that have a sufficiently close labor relationship with the formal sector; specifically, I limit the sample to workers appearing in our sample for at least 5 years between 2000 and 2014.

It is not infrequent that workers in the sample will appear as performing multiple different jobs in the same year. In order to obtain a dataset in which each unit of observation corresponds to a worker and a year, I assign to each worker-year specific pair the location, sector and occupation corresponding to the job that the worker hold for the longest period of time during the corresponding year. However, to determine the total labor income of a worker in a year, I add the labor income earned by the worker in every job in which, according to the data, this worker has been employed in the corresponding year.[5]

[4]Brazilian microregions are groups of municipalities that span the entirety of the Brazilian territory. During our sample period, there were 557 microregions which may themselves grouped into 136 mesoregions.

[5]To compute the total labor income of a worker associated to each job this worker has held, I transform the average monthly earnings reported for each job into a measure of average daily wages,

Figure B.2.1: Migration rates across mesoregions by Year over 2000-2014



B.2.2 Additional Descriptive Statistics on Migration Patterns

In this section, I describe additional statistics about the sample. As illustrated by Figure B.2.1, the overall migration rate increase by about 1.5-2 percentage points between 2000 and 2014. I then describe which demographic groups are the most mobile. Figure B.2.2 illustrates a steep decrease of the migration rate with age. Figure B.2.3 shows that females in the sample are about half as mobile as males. Figure B.2.5 illustrates that the migration rate is increasing for highly educated, decreasing for low-educated.

Table B.2.1 provides descriptive information on the set of workers who appear at least once in the RAIS data between January 2000 and December 2014. The average worker in our sample (weighted by months in sample) is 39.4 years old and has annualized earnings of R$ 18,822 (in 2014 R$), before taxes. Over the sample period, the average annual rate of migration across mesoregions is 7.3%. For males, the average yearly migration rate is 8.1%, whereas it is 4.9% for females. Older workers have a smaller propensity to migrate, with a migration rate of 4.2% for workers between 45 and 64 years old, 7.4% for workers between

and multiply this one by the total number of days between the day of accession and the day of separation into the job reported in the data. If no information on the day of accession or separation is reported, I assume that these ones are January 1 and December 31, respectively.

Figure B.2.2: Migration rates across mesoregions by Year, by age, over 2000-2014



Figure B.2.3: Migration rates across mesoregions by Year, by gender, over 2000-2014



25 and 44. Workers with a high school degree or above have an average migration rate of 8.1%, while it is 5.1% for workers below high school completion.

Figure B.2.4: Migration rates across mesoregions by Year, by race, over 2000-2014

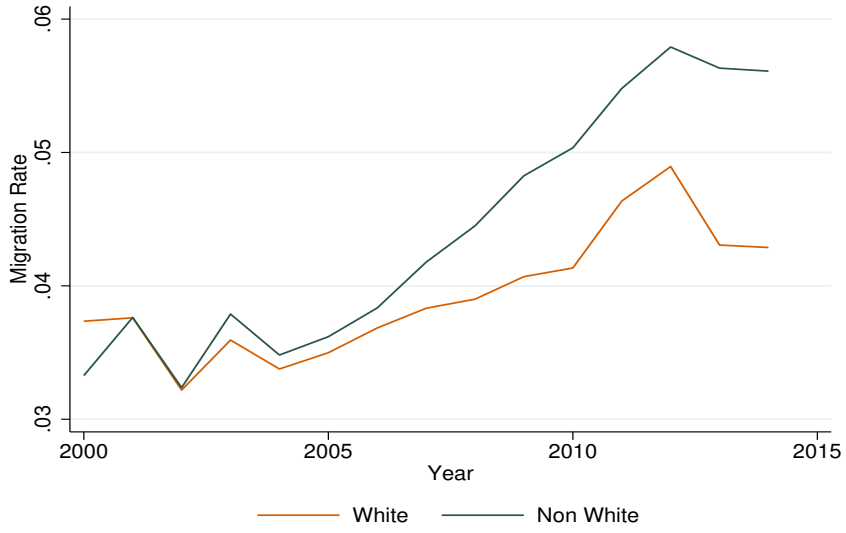


Figure B.2.5: Migration rates across mesoregions by Year, by educational attainment, over 2000-2014



Figure B.2.6: Population Density by Mesoregion - Average over 2000-2014

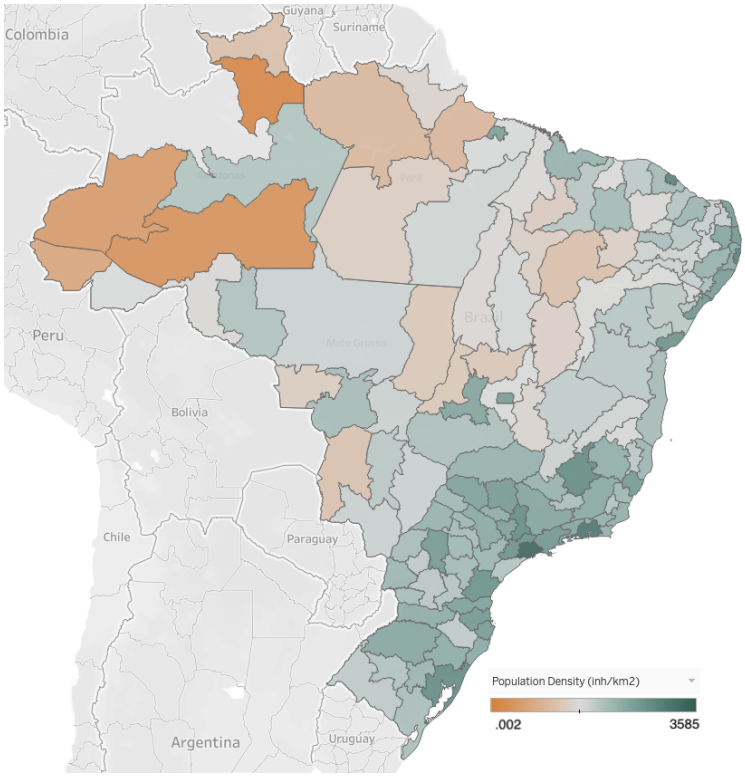


Table B.2.1: Mean and Quantiles of Main Variables

Year	Variable	Mean	Std. dev.	Min	Max	p10	p50	p90
2000	Age	37.22	8.91	25.00	99.00	27.00	36.00	50.00
	Schooling	5.78	2.17	-1.00	11.00	3.00	6.00	9.00
	Tot. Income	1,902	12,790	0.00	871,313	342	1,431	4,746
2014	Age	40.43	10.20	25.00	113.00	28.00	39.00	55.00
	Schooling	6.28	1.86	-1.00	11.00	4.00	7.00	9.00
	Tot. Income	22,976	64,657	0.00	6,068,441	746	5,443	51,529
Avg.	Age	39.40	9.92	25.00	113.00	28.00	38.00	54.00
	Schooling	6.12	1.98	-1.00	11.00	3.00	7.00	9.00
	Tot. Income	18,822	54,226	0.00	6,068,441	425	3,681	45,696

B.3 Appendix to Section 2.4

B.3.1 Algorithm for Simulated Method of Moments

- Parameters to estimate:
  - Information cost (λ_l, λ_h) for low/high internet regions respectively
  - Role of distance (ϕ_0, ϕ_1), so that $\kappa_{jk} = \phi_0 \mathbb{1}_{j \neq k} + \phi_1 dist_{jk}$
  - Decreasing returns in labor in production α
  - Persistence of regional stochastic productivity γ
  - Regional constant productivities $\{A_k\}_{k=1,...J}$
  - Regional constant amenities $\{B_k\}_{k=1,...J}$
- The discount rate is set to $\delta = 0.96$.
- From the wage equation, we can estimate $\{A_k\}_{k=1,...J}$, α and γ:

$$\log w_{kt} = A_k + \theta_{kt} - \alpha \log L_{kt}.$$

  From the OLS regression, treating θ_{kt} as the error term, we get estimates of A_k and α. We can then estimate the persistence parameter γ by fitting a 2-state Markov chain on the sequence of residuals $\hat{\theta}_{kt} = \mathbb{1}\{\theta_{kt} > 0\}$ obtained from the regression. Note that θ_{kt} is correlated with L_{kt} so I will need an instrument to estimate α.

- To estimate the remaining parameters, namely amenities, information cost and the role of distance, we use a simulated method of moments (SMM), based on the gravity equation for migration flows predicted by the model:

$$\Lambda_{jklt} = Q_{jklt} + \frac{1}{\lambda_j} \Delta w_{jkt} - \frac{1}{\lambda_j} K_{jkl} + \frac{1}{\lambda_j} (B_k - B_j) + \frac{\delta}{\lambda_j} \chi_{jkt},$$

  where

$$\Lambda_{jklt} = \log \left(\frac{p_{jkt} p^{\delta}_{klt+1}}{p_{jjt} p^{\delta}_{jlt+1}} \right), \qquad Q_{jklt} = \log \left(\frac{q_{jkt} q^{\delta}_{klt+1}}{q_{jjt} q^{\delta}_{jlt+1}} \right)$$

$$\Delta w_{jkt} = w_{kt} - w_{jt},$$

$$K_{jkl} = \kappa_{jk} + \delta\left(\kappa_{kl} - \kappa_{jl}\right)$$

$$\chi_{jkt} = EV_{kt,t+1} - V_{kt+1} - \left(EV_{jt,t+1} - V_{jt+1}\right)$$

$$EV_{kt,t+1} = \mathbb{E}\left[V_k(\theta_{t+1}, L_t)|\theta_t, L_{t-1}\right]$$

- In the data, I observe Λ_{jklt}, Δw_{jkt} and $dist_{jk}$ for all regions and years, as well as the population distribution $\{L_{jkt}\}_{j,k=1,...J}$. From the wage regression, I also know the current estimated productivity state θ_{kt} in every region-year.

- I employ the following iterative procedure:

  1. Guess a set of parameters $(\lambda_l^0, \lambda_h^0)$, (ϕ_0^0, ϕ_1^0) and $\{B_k^0\}_{k=1,...J}$.

  2. Simulate the model using these parameters and the previously estimated parameters.

  3. For any triplet-year $jklt$, compute the simulated Q_{jklt}^0, K_{jkl}^0 and χ_{jkt}^0, and run the regression:

$$\Lambda_{jklt} - Q_{jklt}^0 = \beta_1\left(\Delta w_{jkt} + \delta\chi_{jkt}\right) + \beta_2 K_{jkl}^0 + D_k - D_j$$

  4. Update $\lambda^1 = \beta_1^{-1}$, $\phi^1 = \beta_2/\beta_1$, $B_k^1 = D_k$, and return to step 1 until convergence.

- Joint estimation:

  – For any $(\lambda_l^0, \lambda_h^0)$, (ϕ_0^0, ϕ_1^0) and $\{B_k^0\}_{k=1,...J}$, define

$$\varepsilon_{jklt} = \Lambda_{jklt} - \left(Q_{jklt} + \frac{1}{\lambda_j}\Delta w_{jkt} - \frac{1}{\lambda_j}K_{jkl} + \frac{1}{\lambda_j}\left(B_k - B_j\right) + \frac{\delta}{\lambda_j}\chi_{jkt}\right).$$

    where Λ_{jklt} and Δw_{jkt} are the observed migration flows and wage gaps, and the remaining terms are simulated using the parameters and the observed states (θ_t, L_{t-1}).

– The model predicts the following moment conditions must hold:

$$\mathbb{E}\left[\varepsilon_{jklt}\right] = 0,$$

$$\mathbb{E}\left[\varepsilon_{jklt} Q_{jklt}\right] = 0,$$

$$\mathbb{E}\left[\varepsilon_{jklt} \Delta w_{jkt}\right] = 0,$$

$$\mathbb{E}\left[\varepsilon_{jklt} K_{jkl}\right] = 0,$$

$$\mathbb{E}\left[\varepsilon_{jklt}\left(B_k - B_j\right)\right] = 0,$$

$$\mathbb{E}\left[\varepsilon_{jklt} \chi_{jkt}\right] = 0,$$

Using these moment conditions, I can then look for the set of parameters (λ_l, λ_h), (ϕ_0, ϕ_1) and $\{B_k\}_{k=1,\ldots J}$ that minimize an objective function constructed from these moment conditions.

- Data collection:
  - For the 100 most populated microregions:
    - Average Wages across all individuals (for now), each year
    - Number of workers per municipality and year
    - Distance between all regions
    - Bilateral annual migration flows between each microregion

B.3.2 Construction of the Instrument for Internet Access

To instrument for the share of residents in a mesoregion with an active internet connection, I rely on the expansion of the internet infrastructure over the time period of the sample. In particular, as documented by Tian (2019), Brazil witnessed a rapid expansion of its internet network during the period between 2008 and 2014. Figure B.3.1 illustrates the change in the fraction of residents with an internet connection between 2000-2007 and 2008-2014 in

each mesoregion. To the map is superimposed the grid of internet infrastructure that was implemented during the 2008-2014 period. These additional elements are backbones of the internet network. These backbone cables are essential parts of the internet infrastructure. From these main cables, smaller ones fan out and provide broadband connections. As discussed by Tian (2019), the strength of the signal decays with distance to the backbone cable, so that providing a connection of satisfactory quality is typically difficult beyond 250 km from the cable.

Motivated by the technical role played by these cables in providing access to internet, I construct, for each mesoregion, a dummy variable equal to 1 if it is on the path of a backbone cable that was added during the period. Since mesoregions can be quite large, it is typically verified that a mesoregion with no backbone cable passing through it will have most of its population centers located at more than 250 km from the backbone cable.

Denoting by Δz_{jt} the indicator variable equal to 1 if mesoregion j is on the path of a backbone cable that was added during the 2008-2014 period, I run the following first stage regression:

$$\Delta int_{jt} = \delta \Delta z_{jt} + \nu_{jt}.$$

As discussed by Akerman et al. (2019), the exclusion restriction at the basis of the identification is that timing of the roll-out of the backbone cables is unrelated to changes in confounding factors that occurred over the same period.

The identifying assumption is that mesoregions close to and farther away from new broadband backbones were on parallel trends in the outcome of interest prior to the completion of the new backbones, and did not experience systematically different idiosyncratic shocks after the new backbones arrived.

As first argued by Tian (2019), there are two main reasons why this assumption is plausible in the context of the internet expansion in Brazil over this period. First, alignment of the backbones was announced at the beginning of the 2008-2014 period and followed other infrastructures that had existed long before 2008, making it harder for policymakers to align the broadband cables in anticipation of economic changes in certain areas. Second, the order

in which municipalities are connected is approximately geographically determined, according to their distances to the submarine cable landing points along the coast. It is thus a priori unlikely that the availability of the new backbones across different municipalities correlates with the temporal variation in the extents of firms' division of labor of areas on and off the new backbone cables in Brazil.

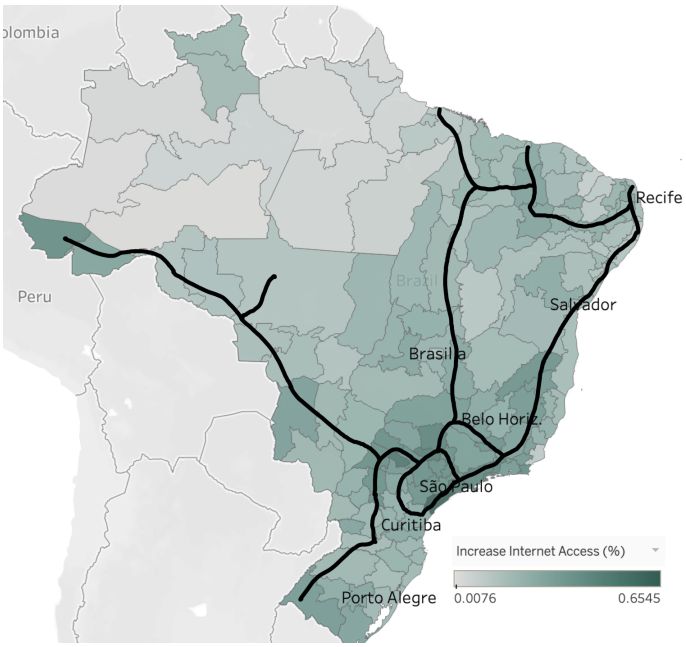
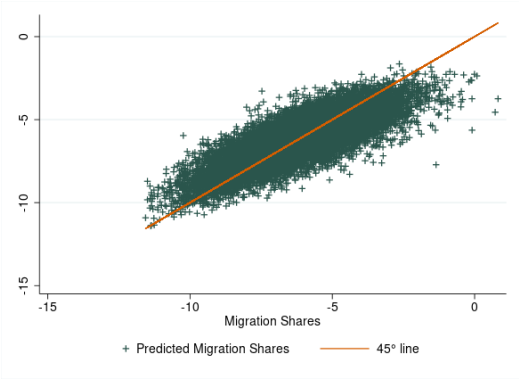


Figure B.3.1: Change in internet access from 2000-2007 to 2008-2014 and new backbone cables

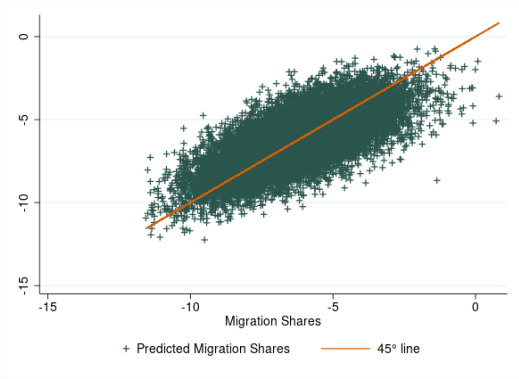
B.4 Appendix to Section 2.5

B.4.1 Fit of Migration Flows

In Figure B.4.1a and Figure B.4.1b, I report the scatter plot of the log of observed migration shares between every pair of mesoregions for all years, against their predicted values according to the model with information frictions, and without, respectively. When information frictions are allowed, the R^2 of the regression of predicted on actual migration costs is 0.70, whereas it is only 0.61 in the model with no information frictions.



(a) Predicted vs. Actual Migration Flows: with Information Frictions



(b) Predicted vs. Actual Migration Flows: No Information Frictions

B.4.2 Heterogeneous Migration Elasticities in the Model with No Information Frictions

B.4.3 List of Local Shocks

In order to empirically identify a mesoregion as having experienced a positive labor market demand shock, I require this mesoregion to verify three criteria. First, there must exist a year such that the average immigration rate in the next four years is at least 50% larger than the average immigration rate in the previous four years; this year is defined as the year of at which the positive labor demand shock too place. Second, the average population in the 4 years preceding the year of the shock is at least 20,000 workers. Third, indexing the year at which the shock took place as t, the total number of immigrants to the shocked region

over the period $t-2$ to $t+4$ must be at least 20,000 workers. The first criterium uses a discontinuity in immigration rates to identify regions experiencing a positive labor demand shock. The second and third criterium restrict the shocks of interest to those taking place in mesoregions that are sufficiently large and that drove a sufficiently large number of workers into the shocked region.

Once I have identified a mesoregion that experienced a positive labor demand shock according to the criteria outlined in the previous paragraph, I focus on the workers that migrated to the shocked region during a period encompassing the two years prior and the four years subsequent to the shock, and relate the timing of their migration decision to different characteristics of both the migrants themselves and of the mesoregion from which they migrated.

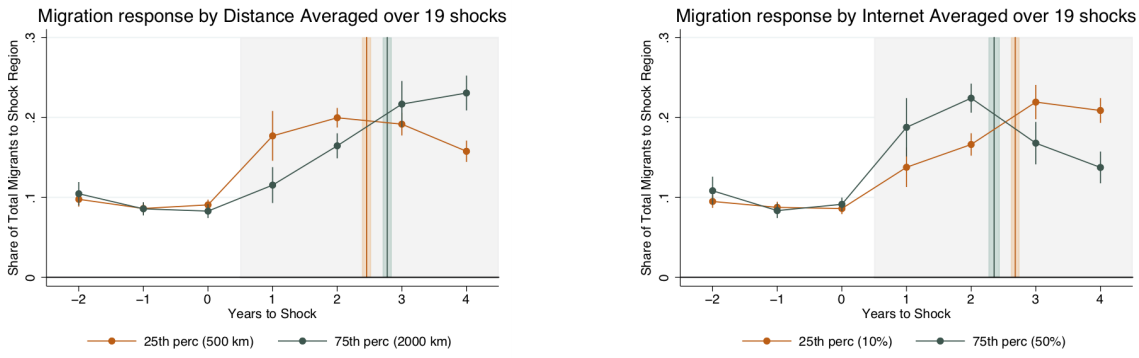
Given that the last year for which I observe migration decisions in the data is 2014 and that I want to study the evolution of the migration flows to a region experiencing a positive labor demand shock in the four years subsequent to the shock, I only search for regions that experienced positive labor demand shocks prior to 2010. This procedure to identify such regions identified a total of ten mesoregions which experience positive local labor demand shocks. As summarized in Table B.4.1, these are diverse in their location, the year in which they took place, the total number of workers they draw into the shocked regions and the underlying cause of the shock.

Table B.4.1: Description of Positive Local Labor Demand Shocks

Municipality	State	Year Shock	Number Migrants	Source Shock
Ipojuca	Pernambuco	2009	89,067	Refinery Construction
Natal	R. G. do Norte	2008	74,294	Oil Boom
Santo Antonio	Rondonia	2008	65,746	Dam Construction
Maceio	Alagoas	2008	49,943	Tourism Boom
Belo Monte	Para	2010	40,644	Dam Construction
Uberaba	Minas Gerais	2008	40,301	Sugar Cane Energy
Araucaria	Parana	2006	37,684	Tourism Boom
Cidelandia	Maranhao	2010	37,216	Palm Oil Boom
Suape	Pernambuco	2008	32,752	Refinery Construction
Itabira	Minas Gerais	2010	21,115	Mining Boom

B.4.4 Replicating Local Shocks in the Model

Figure B.4.2: Average Delay in the Migration Response to the 10 Local Shocks



(a) Distance: Data

(b) Internet: Data

Figure B.4.2a and Figure B.4.2b illustrate the estimates that arise from pooling the data across the ten shock listed in Table B.4.1. They illustrate, for each year between $t-2$ and

$t+4$, the predicted migration probability when the corresponding covariate X^k_{it+s} is set to its 25% percentile (labeled as "Low" and painted in orange) or to its 75% percentile (labeled as "High" and painted in green) are set to their mean values. The whiskers attached to each dot represent the 95% confidence interval for each predicted migration probability. The dark thin vertical lines indicate the estimated expected number of years of delay implied by the expected probabilities, and the light-colored thick vertical lines illustrate the corresponding 95% confidence intervals. The standard errors of the estimates of the implied expected probabilities and the expected number of years of delay are computed using the Delta method and standard errors for the estimates of the parameter $\{\gamma^k_{it+s}; k = 1, \ldots, K\}$ clustered by year and mesoregion of origin of the migrant (i.e. by the mesoregion in which the migrant was located at period $t+s-1$).

The results in Figures B.4.2a and B.4.2b illustrate that everything else equal, workers living in geographically close mesoregions or in mesoregions in higher broadband internet penetration tend to react faster to positive labor demand shocks happening in mesoregions other than their location of residence. I interpret these estimates as suggestive of the hypotheses that, everything else equal: (a) workers tend to have better information about labor demand shocks taking place in markets that are geographically close to their location of residence; and (b) workers located in areas with higher internet penetration have better information about every labor demand shocks, no matter where this one took place.

B.5 Additional Figures

B.5.1 Cardell Distribution

$$g_\beta(z) = \frac{1}{\beta} \sum_{n=0}^{\infty} \frac{(-1)^n \, e^{-nz}}{n! \Gamma(-\beta n)}.$$

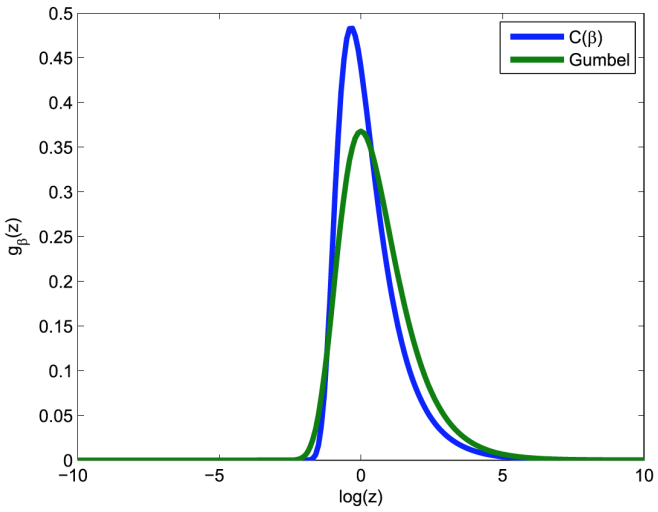


Figure B.5.1: Distribution Cardell with dispersion $\beta = 0.5$, vs. Gumbel. Source: Dasgupta and Mondria (2014).

Appendix C

Appendix for "The Role of Establishment Size in the City-Size Premium in Spain"

C.1 Data Appendix

In this appendix, we explore the robustness of our main results. In the first section of the appendix we present some additional evidence on the correlation between establishment size, and city size as well as establishment size and city earnings. In the second section, we replicate the main results on the city-size earnings premium for establishment size dummies and firm size dummies.

C.1.1 Data Definitions

Urban Area. The most disaggregated geographical level at which we observe people's location is a municipality. There are 8,131 municipalities in Spain, with an average area of 62.2 km^2 and an average population of 5,746 inhabitants per municipality. We group municipalities into urban areas using the official definition for urban areas published by Spain's Ministry of Housing in 2008. We exclude urban areas with less than 40,000 inhabitants because in these cases a workplace municipality is not provided for anonymity concerns. Moreover, the sample does not include the autonomous cities of Ceuta and Melilla, and urban areas in the Basque Country and Navarre because they have an independent social security administration. The final sample includes 76 urban areas, 161,474 persons, and 7,727,861 person-month observations. We employ the measure of city size calculated by De la Roca and Puga (2017). They measure the size of an urban area by computing the number of people within 10 km of each inhabitant in the urban area and taking the average. The result is the number of people within 10 km of the average person. This measure is intended to capture the fact that for an individual the relevant size of a city may not only depend on overall population, but also on how spatially concentrated is the population. As showed in De la Roca and Puga (2017), the measures of density are very highly correlated with total population counts. In order to calculate this measure of city size, it is necessary to have a detailed grid of the population in Spain. De la Roca and Puga (2017) use the 1km by 1km population grid that Goerlich and Cantarino (2013) calculated for 2006, and we use the density in 2006 as a time-invariant measure of city size.

Definition of an establishment. To define establishments, we use the Código de Cuenta de Cotización (CCC), a tax identifier used for contributions to Social Security that is associated with an employer-province pair. If a firm has many plants in a province, these will usually be grouped under the same CCC. Thus, our definition of establishment does not necessarily identify each physical plant location of the firm, but gives a disaggregation of the firm at the province level. A potential drawback of this measure, is that in some cases a firm may have more than one CCC in a province as required for administration purposes. In practice, only 1 percent of the city-firm pairs in the sample have more than one CCC. Throughout the paper, we refer to the employer-province pair as establishment to distinguish it from the firm, which spans all provinces.

Consequently, the number of co-workers refers to the number of workers associated with a CCC and not a physical plant. Ideally, we would have information on the number of co-workers within the same urban area. However, since the CCC identifies a firm-province pair, it includes co-workers at plants located in different urban areas of the same province. In practice, this is a limited concern since provinces are small and most have only one urban area. It is worth noting that the location of the worker is precisely known. But a worker in an urban area will be assigned as co-workers all the individuals working for the same firm in the whole province.[1].

Finally, although we observe the number of co-workers for each worker in the sample, we do not observe any characteristics of these co-workers. Only 26% of the establishments have more than one worker in the sample. As a result, although we have a good measure of the number of workers in a firm, we do not observe enough of them to estimate firm fixed effects.

Employer Tax ID. On top of the CCC we also observe the tax identifier of the employer to which all the CCC belong. We can therefore observe whether a worker works for a firm with multiple establishments in different provinces and the total number of workers employed

[1]There is a total of 50 provinces in Spain but due to regional differences in tax collection autonomy we do not have data for four provinces: Bizkaia, Gipuzkoa, Araba, and Navarra. The final sample includes 46 provinces. Most provinces contain only one urban area. The province of Alicante has 7 urban areas, 8 other provinces have 3 urban areas and 10 have 2 urban areas.

at that firm. In Section C.1.3 of the Appendix, we test whether the establishment or the firm co-workers seem more relevant for determining the co-worker earnings premium. We find that firm co-workers is no longer associated with higher earnings when controlling for establishment co-workers, and therefore we continue our analysis using establishment co-workers. We check the robustness of the results looking at firm co-workers instead of establishment co-workers in Section C.1.3 of the Appendix.

Experience. We are particularly interested in constructing variables capturing aspects of accumulated experience in an urban area. In order to construct urban area variables, we aggregate workplace municipalities to the definition of an urban area. We use the information to construct variables on experience accumulated in different city-size categories following the method of De la Roca and Puga (2017).

Earnings. In order to obtain monthly uncapped labor earnings, we combine information from two sources. First, the social security records include monthly top-coded earnings that exclude payment of over-time hours and other in-kind payments. Second, the tax administration records include all monetary and in-kind labor earnings in a given year and is not top-coded. We allocate the labor earnings from the tax administration across months according to the fraction of top-coded earnings from the social security records that was earned in each month. Earnings are deflated using the Consumer Price Index and converted into 2009 euros, and are then adjusted using the part-time coefficient to measure the full-time equivalent real daily earnings.

C.1.2 Summary statistics

This section, presents summary statistics on the dependent variables and the controls. TableC.1.1 includes the mean, standard deviation, minimum and maximum. The sample is relatively young, with low tenure and low job experience. We suspect this is the result of restricting the sample to exclude individuals in agriculture, public administration, education, and healthcare. It could also be in part due to excluding individuals from the Basque Country and Navarre, since these are regions with high levels of education and labor market attachment compared to the rest of Spain.

Table C.1.1: Summary Statistics

Variable	Units	Mean	Std. Dev.	Min	Max
Daily earnings	Euros	72.51	90.45	0.00	37,042
Number of co-workers	Persons	575.34	1,859.05	0.00	22,077
Age	Years	34.39	7.07	17.08	47.92
Less than high school education	Dummy	0.50	0.460.50	0.00	1.00
High school education	Dummy	0.31	0.46	0.00	1.00
University and above education	Dummy	0.19	0.39	0.00	1.00
Days worked in a month	Days	29.86	3.30	1.00	31.00
Tenure	Years	4.74	4.97	0.00	32.59
Exper.	Years	11.23	6.48	0.00	36.59
Exper. in 1st-2nd largest cities	Years	4.39	6.45	0.00	31.43
Exper. in 3rd-5th largest cities	Years	1.03	3.42	0.00	31.27
Exper. outside 5 largest cities	Years	4.29	6.07	0.00	33.45
Exper. in current city	Years	9.94	6.36	0.00	33.42
Fixed Contract	Dummy	0.28	0.45	0.00	1.00
Part-time Contract	Dummy	0.09	0.28	0.00	1.00
Very-high-skilled occupation	Dummy	0.08	0.27	0.00	1.00
High-skilled occupation	Dummy	0.12	0.32	0.00	1.00
Middle-skilled occupation	Dummy	0.23	0.42	0.00	1.00
Low-skilled occupation	Dummy	0.46	0.50	0.00	1.00
Very-low-skilled occupation	Dummy	0.11	0.31	0.00	1.00
Observations	7,019,750				

To test how representative is our sample, we compare it to the EU Labor Force Survey (LFS) conducted by Eurostat. We select the LFS sample to be as similar as ours as possible, given the available characteristics. We keep only males, from 15 to 75, born in Spain,

who are not employed in the primary sector. Our sample is a bit younger, less educated and with less tenure. However, it is worth noting that the tenure variable in the LFS measures the number of years since first employed by the current employer. In contrast, our measure of tenure counts the number of years continuously employed but the current employer. These two measures could differ if there is discontinuous involvement with the same employer. Moreover, the low education could also be due to under-reporting in the labor contracts. There is low monitoring and incentives to report the education attainment correctly. However, since we include worker fixed effects for the main results, the possible under-reporting of education is unlikely to affect our estimates.

A brief comment on the experience value, which can seem a bit low, is also in place. The average age in the sample is around 34.5, and the average experience is 11.3. Given that 60 percent of the population didn't finish college, this implies an average of 5 years of non-employment since finishing formal education. To put this number in perspective, it is important to keep in mind that Spain is a country with high unemployment levels, specially among young workers. Youth unemployment in Spain was 40 percent in 1996. It decreased steadily up to 2008, when it reached 20 percent. After the recession, it shot up to its 55 percent peak around 2013, when it finally started decreasing again. In 2010, youth unemployment reached 30 percent, still far from the pre-2008 levels. In this context, five years of not-employment does not seem excessive or problematic.

C.1.3 Firm and Establishment Co-workers

This section explores whether establishment size or firm size seems more relevant to estimate the city size co-worker premium and the co-worker earnings premium. To examine this, we first regress the log of the number of co-workers at the establishment level for a worker i on city fixed effects γ_c and the log of co-workers at the firm, over all its establishments.

$$\ln(\text{establishment coworkers}_i) = \alpha_n + \eta_t + \gamma_c + \beta \ln(\text{firm coworkers}_i) + \varepsilon_i$$

Table C.1.2: City Size Co-worker Premium: Establishments vs Firms

	(1) Ln Estab. Co-Workers	(2) City FE	(3) Ln Firm Co-Workers	(4) City FE
Ln City Size		0.110**		-0.050
Ln Estab. Co-Workers			1.151***	
Ln Firm Co-Workers	0.671***			
Time and Sector FE	Y		Y	
Observations	7,019,750	76	7,019,750	76
R^2	0.836	0.085	0.8312	0.0098

Note: p<0.1, ** p<0.05, ***p<0.001. Column (1) reports the results from regressing log establishment workers on city, time, and sector fixed effects. Column (2) reports the results from regressing the city fixed effects from Column (1) on log city size. Column (3) reports the results from regressing log firm workers on city, time, and sector fixed effects, Column (4) reports the results from regressing the city fixed effects from Column (3) on log city size. City size is measured as the number of people within 10km of the average person, an establishment corresponds to a firm-province tax identifier. Data Source: Spain's Continuous Sample of Employment Histories 2006-2013.

We then regress the city fixed effects on log city size

$$\gamma_c = \alpha + \theta \ln(\text{city size}_c) + \varepsilon_c$$

We interpret θ as the city size co-worker premium, controlling for firm size. Column 2 of table C.1.2 presents the estimated elasticity of establishment co-workers with respect to city size, controlling for firm co-workers, as well as the coefficient, β, on log firm co-workers. Given firm co-workers, workers in a city twice as large have on average 11 percent more co-workers in their establishment. Column 4 of table C.1.2 reports the city size co-worker premium at the firm, controlling for establishment co-workers. Given a number of co-workers in the establishment, there is no further correlation between city size and the number of co-workers at the firm level. For this reason, in the main analysis we use the establishment co-workers to measure the city size co-worker premium. However, we do not control for the number of co-workers at the firm. Since these two are correlated, the effect of changing the number of co-workers at the establishment includes as well the change in co-workers at the firm level.

Table C.1.3: Co-worker Earnings Premium: Establishments vs Firms

	(1) Ln Earn.	(2) Ln Earn.	(3) Ln Earn.
Ln Estab. Co-Workers	0.055***	0.021***	0.022***
Ln Firm Co-Workers	0.022***	0.026***	0.004***
Worker FE	N	N	Y
Controls	N	Y	Y
Observations	7,019,750	7,019,750	7,019,750
R^2	0.111	0.504	0.063

Note: p<0.1, ** p<0.05, ***p<0.001. This table reports the results from regressing log earnings on city fixed effects and a series of controls, and then regressing the city fixed effects on log city size. City size is the number of people living within 10km of the average person. Controls include education, occupational categories, industry and time fixed effects, and type of contract. Data Source: Spain's Continuous Sample of Employment Histories 2006-2013.

Next, we estimate the co-worker earnings premium controlling for the number of co-worker at the firm level and present the results in Table C.1.3. There exists both a firm and an establishment co-worker earnings premium. However, once we include both firm and establishment co-workers with controls and worker fixed effects the establishment co-worker earnings premium is about five times larger than the firm co-worker earnings premium. This suggests that, although both naturally move together, the number of co-workers at the establishment level is a better predictor of earnings.

C.2 Additional Results

C.2.1 A Closer Look into Mover Regressions

Although the inclusion of worker fixed effects allows us to control for unobservable characteristics, it also requires a sample selection of movers only. Whenever we include worker fixed effects, our identification of city fixed effects is coming from workers who move across cities.

There are several reasons why the estimates might change as a result of including worker fixed effects. First, as intended, we could be controlling for unobserved characteristics of the worker such as innate ability. Second, since we are selecting workers based on whether they moved or not, it may be that workers who move are not representative of the whole population and they may not experience the same earnings premium as non-movers. If workers only move in response to an attractive job offer, the increase in earnings as a result of the move may be higher than what we would observe if any person were to move at random to a large city. If on the contrary, workers move for other reasons and need some time to find a good job offer, the increase in earnings may be lower than the average difference in earnings across cities. There could also be learning effects as emphasized in De la Roca and Puga (2017) so that as workers stay in a large city, they accumulate valuable experience that leads to faster growth in earnings relative to a smaller city.

First, we test whether movers are representative of the whole population. We compute the city-size earnings premium for movers by running the cross-sectional regression not including worker fixed effects, but only for those workers that move at some point in our sample. The results are included in Columns 2 of Table C.2.1 while Column 1 presents the comparable results for the full sample. Reducing the sample to only movers has a small positive effect on the cross-sectional coefficient. For movers, the elasticity between earnings and city size is .049 percent as compared to .046 percent in the whole population. Therefore, it doesn't seem that workers who move at some point are much different than the general population.

Table C.2.1: Selection of Movers and Worker Fixed Effects

	(1)	(2)
	Ln Earnings City FEs from Step 1	
	Full Sample	Movers Only
Ln City Size	0.046***	0.049***
Worker Fixed Effects	N	N
Controls	Y	Y
Observations	7,727,861	1,094,338

Note: * p<0.1; ** p<0.05, ***p<0.01. This table reports the results from regressing log earnings and log co-workers on city fixed effects and a series of controls, and then regressing the city fixed effects on log city size selecting only movers. City size is the number of people living within 10km of the average person. Controls include education, occupational categories, industry and time fixed effects, and type of contract. Data Source: Spain's Continuous Sample of Employment Histories 2006-2013.

Finally, we look a little deeper into moves across cities an how they relate to moves across firm sizes. In Table C.2.2, we look at the transition matrix between firm size categories for workers who move between small and large cities. We compare the transition matrix between firm size categories observed in the data to what would be expected if the moves across firm size categories were random, given the differences in firm size distributions across city sizes. We find that, relative to random moves, workers tend to move within a firm size category (the diagonal elements of the matrix are slightly larger in the data than if the moves were random). This indicates that there is some degree of sorting of workers to firm sizes. Workers who work at a large firm and move from a small to large city are more likely to work for a large firm when they arrive in the big city. However, we do not necessarily see evidence of selection. For example, it does not seem that workers who are employed at a small firm are moving to the big city *because* they receive a job offer from a large firm. If anything they are slightly less likely to work at a large firm than if they had received a random job offer.

Table C.2.2: Non-random Moves across Firm Size Categories

	Moving to:	Large Firm	Middle Firm	Small Firm
		Random		
Moving from:	Large Firm	23%	9%	11%
	Middle Firm	12%	4%	6%
	Small Firm	19%	7%	9%
		Data		
Moving from:	Large Firm	35%	8%	8%
	Middle Firm	7%	6%	4%
	Small Firm	14%	7%	11%

Note: this table reports the percentage of people who move across city sizes that also move across firm sizes and compares it to the percentages if the moves across firm sizes were random. A large city is defined as one of the three largest cities, that is, Madrid, Barcelona or Valencia, and a small city refers to a not city that is note in the top 5 largest cities. A small firm is a firm with fewer than 20 employees, a medium firm is a firm that has between 20 and 100 employees and a large firm is a firm with more than 100 employees. Data Source: Spain's Continuous Sample of Employment Histories 2006-2013.

C.2.2 Heterogeneity in the Co-worker Earnings Premium Across Cities

Moving to an establishment with more co-workers is associated with large earnings gains. However, it is possible that these gains are not the same across all cities. We calculate the within-city co-worker earnings premium, namely the expected premium from increasing the number of co-workers conditional on staying in the same city, and we show that it is decreasing in city size. This novel fact has important implications. In decomposing the city-size earnings premium , our goal is to understand what share of the city-size premium is driven by a contemporaneous move to a larger establishment. The fact that the within-city co-worker earnings premium is smaller in large cities means that the expected effect of moving from a small establishment in a small city to a large establishment in a large city will be smaller than the naive estimate of combining the average co-worker earnings premium and the average city-size co-worker premium. Second, the differences across cities in the co-worker earnings premium has important implications for the dynamic earnings gains from living in a large city. Climbing the establishment-size ladder will have smaller effects in a

large city suggesting that the dynamic gains from learning documented by De la Roca and Puga (2017) are less likely to manifest by faster transitions to larger establishments. We discuss the dynamic implications further in section 3.4.2.

A final implication of the fact that the co-worker earnings premium is smaller in large cities is that it provides supporting evidence that firms in large cities have less monopsony power than firms in small cities. This is consistent with the theory of firm sorting proposed by Manning (2010). In his theory, large firms sort to large cities because they face a higher elasticity of labor supply in large cities, i.e. they have less monopsony power. This means that in order to grow large, they do not need to raise wages as much in a large city relative to a small city. Thus, the relationship between establishment size and earnings is weaker in large cities.

In order to document the relationship between the within-city co-worker premium and city size, we again follow a two-step procedure. First, we run a regression similar to equation 3.1, but this time allowing the coefficient on log number of co-workers to vary across cities. Specifically, we run the regression

$$\ln(w_{it}) = \alpha_{c(i,t)} + \delta_t + \mu_i + \beta_{c(i,t)} \ln(coworkers_{it}) + x_{it}'\psi + \varepsilon_{it}, \tag{C.2.1}$$

where c(i,t) is the city where individual i lives at time t, $\alpha_{c(i,t)}$ is the corresponding city fixed effect, δ_t is a month-year fixed effect, μ_i is an individual fixed effect, x_{it} is a vector of characteristics of worker i at time t which include education, occupation, sector, experience, tenure, and type of contract; and ε_{it} is the residual earnings. Next, we regress the coefficient on log co-workers, β_c, on log city size

$$\beta_c = \gamma \ln(city\ size_c) + u_c. \tag{C.2.2}$$

We report the results of these regressions in table C.2.3. Columns 2 and 4 give the relationship between the within-city co-worker earnings premium and city size. Without controlling for worker fixed effects, the relationship is negative but statistically insignificant at the 10% level. After adding worker fixed effects, the relationship becomes significant

at the 1% level. Controlling for worker fixed effects, the within-city co-worker earnings premium is identified off of workers who move across establishment sizes within a city thus controlling for the possibility that within cities, workers of different abilities are sorting to different establishments. The fact that the estimate of the within-city co-worker premium declines more for large cities once we control for worker fixed effects suggests that sorting plays a larger role in large cities. This is consistent with the evidence presented by Dauth et al. (2019) suggesting that sorting of high skilled individuals to large establishments is more important in large cities.

The within-city co-worker earnings premium is 0.37 ($= 0.54 \times \ln(2)$ percentage pointss-maller in cities twice as large. In practice, this corresponds to a significant difference in the co-worker earnings premium between large and small cities. In Utrera, the smallest city in our sample, the elasticity between earnings and city size is .053 percent, while in Madrid the elasticity is .021 percent, less than half the size of the elasticity in Utrera. Figure 3.1 presents the relationship between the within-city co-worker earnings premium and city size graphically. Note that while there is a wide range of these premia within small cities, for large cities, the within-city co-worker premium hovers around 2.5 percent, similar to the economy wide average estimated in panel 2, column 3 of table 3.1.

Table C.2.3: Co-workers Earnings Premium and City Size

	Ln. Earn	β_c	Ln Earn.	β_c
	(1)	(3)	(4)	(6)
Ln City Size		-0.0031		-0.0054
		(0.0023)		(0.0020)***
City FE	Y		Y	
Worker Fixed Effects	N		Y	
Controls	N		Y	
N	7,019,750	76	7,019,750	76
R^2	0.5160	0.0210	0.387	0.0338

Note: * p<0.1; ** p<0.05, ***p<0.01. This table reports the results from regressing log earnings on log establishment workers. Controls include experience, tenure, education, occupational categories, time and sector fixed effects, and dummies for part time and fixed term contracts. Data Source: An establishment corresponds to a firm-province tax identifier. Clustering is implemented at the person level. Data Source: Spain's Continuous Sample of Employment Histories 2006-2013.

Bibliography

Acemoglu, Daron and Pascual Restrepo, "Robots and Jobs," *mimeo*, 2017.

Adão, Rodrigo, "Worker Heterogeneity, Wage Inequality, and International Trade: Theory and Evidence from Brazil," *mimeo*, November 2016.

_ **, Costas Arkolakis, and Federico Esposito**, "Spatial Linkages, Global Shocks, and Local Labor Markets: Theory and Evidence," *mimeo*, February 2019.

Ahlfeldt, Gabriel M., Stephen J. Redding, Daniel M. Sturm, and Wolf Nikolaus, "The Economics of Density: Evidence from the Berlin Wall," *Econometrica*, 2015, *83* (6), 2127–2189.

Ahn, Hyungtaik and Charles F. Manski, "Distribution Theory for the Analysis of Binary Choice under Uncertainty with Nonparametric Estimation of Expectations," *Journal of Econometrics*, 1993, *56*, 291–321.

Aker, Jenny C., "Information from Markets Near and Far: Mobile Phones and Agricultural Markets in Niger," *American Economic Journal: Applied Economics*, July 2010, *2* (3), 46–59.

_ **, Paul Collier, and Pedro C. Vicente**, "Is Information Power? Using Mobile Phones and Free Newspapers during an Election in Mozambique," *The Review of Economics and Statistics*, 2017, *99* (2), 185–200.

Akerman, Anders, Edwin Leuven, and Magne Mogstad, "Information Frictions, Internet and the Relationship between Distance and Trade," *Working Paper*, 2019.

Allen, Treb, "Information Frictions in Trade," *Econometrica*, 2014, *82* (6), 2041–2083.

Allende, Claudia, Francisco Gallego, and Christopher Neilson, "Approximating the Equilibrium Effects of Informed School Choice," *Princeton Industrial Relations Sections Working Paper*, 2019.

Armantier, Olivier, Wändi Bruine de Bruin, Simon Potter, Giorgio Topa, Wilbert van der Klaauw, and Basit Zafar, "Measuring Inflation Expectations," *Annual Review of Economics*, 2013, *5* (1), 273–301.

Artuç, Erhan and John McLaren, "Trade Policy and Wage Inequality: A Structural Analysis with Occupational and Sectoral Mobility," *Journal of International Economics*, 2015, *97* (2), 278–294.

_ , **Shubham Chaudhuri, and John McLaren**, "Trade Shocks and Labour Adjustment: A Structural Empirical Approach," *American Economic Review*, 2010, *100* (3), 1008–1045.

Autor, David H. and David Dorn, "The Growth of Low-Skill Service Jobs and the Polarization of the US Labor Market," *American Economic Review*, August 2013, *103* (5), 1553–97.

_ , _ , **and Gordon H. Hanson**, "The China Syndrome: Local Labor Market Effects of Import Competition in the United States," *American Economic Review*, 2013, *103* (6), 2121–2168.

Bah, Tijan and Catia Batista, "Understanding the Willingness to Migrate Illegally: Evidence from a Lab in the Field Experiment," *NOVAFRICA Working Paper*, 2018, *1803*.

Bartik, Alexander W., Janet Currie, Michael Greenstone, and Christopher R. Knittel, "The Local Economic and Welfare Consequences of Hydraulic Fracturing," *American Economic Journal: Applied Economics*, October 2019, *11* (4), 105–55.

Bartik, Timothy J., *Who Benefits from State and Local Economic Development Policies?*, Kalamazoo, MI: W.E. Upjohn Institute for Employment Research, 1991.

Baseler, Travis, "Intrahousehold Spillovers in the Migration Decision: Evidence from Thailand's Regional Cities Project," *mimeo*, December 2016.

_ , "Hidden Income and the Perceived Returns to Migration: Experimental Evidence from Kenya," *mimeo*, January 2019.

Bayard, Kimberly and Kenneth R. Troske, "Examining the Employer-Size Wage Premium in the Manufacturing, Retail Trade, and Service Industries Using Employer-Employee Matched Data," *American Economic Review*, May 1999, *89* (2), 99–103.

Bayer, Patrick, StephenÂăL. Ross, and Giorgio Topa, "Place of Work and Place of Residence: Informal Hiring Networks and Labor Market Outcomes," *Journal of Political Economy*, 2008, *116* (6), 1150–1196.

Bayer, Patrik, Robert McMillan, Alvin Murphy, and Christopher Timmins, "A Dynamic Model of Demand for Houses and Neighborhoods," *Econometrica*, 2016, *84* (3), 893–942.

Beaudry, Paul, Mark Doms, and Ethan Lewis, "Should the Personal Computer be Considered a Technological Revolution? Evidence from U.S. Metropolitan Areas," *Journal of Political Economy*, 2010, *118* (5), 988–1036.

Behrens, Kristian, Gilles Duranton, and Frédéric Robert-Nicoud, "Productive Cities: Sorting, Selection, and Agglomeration," *Journal of Political Economy*, 2014, *122* (3), 507–553.

BenYishay, Ariel and A Mushfiq Mobarak, "Social Learning and Incentives for Experimentation and Communication," *The Review of Economic Studies*, 07 2018, *86* (3), 976–1009.

Beraja, Martin, Erik Hurst, and Juan Ospina, "The Aggregate Implications of Regional Business Cycles," *mimeo*, February 2019.

Bergman, Peter, Eric W Chan, and Adam Kapor, "Housing Search Frictions: Evidence from Detailed Search Data and a Field Experiment," Working Paper 27209, National Bureau of Economic Research May 2020.

_ , **Raj Chetty, Stefanie DeLuca, Nathaniel Hendren, Lawrence F Katz, and Christopher Palmer**, "Creating Moves to Opportunity: Experimental Evidence on Barriers to Neighborhood Choice," Working Paper 26164, National Bureau of Economic Research August 2019.

Berman, Nicolas, Vincent Rebeyrol, and Vincent Vicard, "Demand Learning and Firm Dynamics: Evidence from Exporters," *The Review of Economics and Statistics*, 2019, *101* (1), 91–106.

Bertoli, Simone, Jesús Fernández-Huertas Moraga, and Lucas Guichard, "Rational Inattention and Migration Decisions," *Working Paper*, 2019.

Bilbiie, Florian, Fabio Ghironi, and Marc Melitz, "Endogenous Entry, Product Variety and Business Cycles," *Journal of Political Economy*, 2012, *120* (2), 304–345.

Bishop, Kelly C. and Alvin D. Murphy, "Estimating the Willingness to Pay to Avoid Violent Crime: A Dynamic Approach," *American Economic Review: Papers & Proceedings*, 2011, *101* (3), 625–629.

Blanchard, Olivier and Lawrence F Katz, "Regional Evolutions, Brooking Papers on Economic Activity," *Economic Studies Program, The Brookings Institution*, 1992, *23* (1), 1–75.

Bloom, Nicholas, Fatih Guvenen, Ben Smith, Jae Song, and Till M. von Wachter, "Inequality and the Disappearing Large Firm Pay Premium," 2019.

Borjas, George J., "Self-Selection and the Earnings of Immigrants," *American Economic Review*, 1987, *77* (4), 531–553.

_ , "The Economic Analysis of Immigration," in Orley C. Ashenfelter and David Card, eds., *Handbook of Labor Economics*, Amsterdam: Elsevier North-Holland, 1999, pp. 1697–1760.

Brinkman, Jeffrey, Daniele Coen-Pirani, and Rui Castro, "Estimating a Dynamic Equilibrium Model of Firm Location Choices in an Urban Economy," Working Papers 12-26, Federal Reserve Bank of Philadelphia 2012.

Brown, Charles and James Medoff, "The Employer Size-Wage Effect," *Journal of Political Economy*, 1989, *97* (5), 1027–59.

Brown, Jeffrey R. and Austan Goolsbee, "Does the Internet Make Markets More Competitive? Evidence from the Life Insurance Industry," *Journal of Political Economy*, 2002, *110* (3), 481–507.

Brown, Zach Y. and Jihye Jeon, "Endogenous Information Acquisition and Insurance Choice," *Working Paper*, 2019.

Bruin, Wändi Bruine De, Charles F. Manski, Giorgio Topa, and Wilbert van der Klaauw, "Measuring consumer uncertainty about future inflation," *Journal of Applied Econometrics*, 2011, *26* (3), 454–478.

Bryan, Gharad and Melanie Morten, "The Aggregate Productivity Effects of Internal Migration: Evidence from Indonesia," *Journal of Political Economy*, 2018, *forthcoming.*

_ , Shyamal Chowdhury, and Ahmed Mushfiq Mobarak, "Under-investment in a Profitable Technology: The Case of Seasonal Migration in Bangladesh," *Econometrica*, 2014, *82* (5), 1671–1748.

Burdett, Kenneth and Dale T. Mortensen, "Wage Differentials, Employer Size, and Unemployment," *International Economic Review*, May 1998, *39* (2), 257–273.

Busso, Matias, Jesse Gregory, and Patrick Kline, "Assessing the Incidence and Efficiency of a Prominent Place Based Policy," *American Economic Review*, 2013, *103* (2), 897–947.

Bustos, Paula, Bruno Caprettini, and Jacopo Ponticelli, "Agricultural Productivity and Structural Transformation. Evidence from Brazil," *American Economic Review*, 2016, *106* (6), 1320–1365.

Caldwell, Sydnee and Nikolaj Harmon, "Outside Options, Bargaining, and Wages: Evidence from Coworker Networks," *Working Paper*, January 2019.

Caliendo, Lorenzo, Luca David Opromolla, Fernando Parro, and Alessandro Sforza, "Goods and Factor Market Integration: A Quantitative Assessment of the EU Enlargement," *mimeo*, February 2018.

_ , Maximiliano Dvorkin, and Fernando Parro, "Trade and Labor Market Dynamics: General Equilibrium Analysis of the China Shock," *Econometrica*, 2019, *forthcoming.*

Caplin, Andrew and Mark Dean, "Revealed Preference, Rational Inattention, and Costly Information Acquisition," *American Economic Review*, July 2015, *105* (7).

_ , _ , and John Leahy, "Rationally Inattentive Behavior: Characterizing and Generalizing Shannon Entropy," *Working Paper*, February 2019.

Combes, Pierre-Philippe and Laurent Gobillon, "The Empirics of Agglomeration," *Handbook of Regional and Urban Economics*, 2015, *5*, 247–348.

_ , Gilles Duranton, and Laurent Gobillon, "Spatial Wage Disparities: Sorting Matters!," *Journal of Urban Economics*, 2008, *63* (2), 723–742.

_ , _ , _ , Diego Puga, and Sébastian Roux, "The Productivity Advantage of Large Cities: Distinguishing Agglomeration from Firm Selection," *Econometrica*, 2012, *80* (6), 2543–2594.

Cunha, Flavio and James J Heckman, "Identifying and Estimating the Distributions of Ex Post and Ex Ante Returns to Schooling," *Labour Economics*, 2007, *14*, 870–893.

Dahl, Gordon B., "Mobility and the Return to Education: Testing a Roy Model with Multiple Markets," *Econometrica*, 2002, *70* (6), 2367–2420.

Dasgupta, Kunal and Jordi Mondria, "Inattentive Importers," *mimeo*, August 2014.

Dauth, Wolfgang, Sebastian Findeisen, Enrico Moretti, and Jens Suedekum, "Matching in Cities," NBER Working Papers 25227, National Bureau of Economic Research, Inc November 2019.

Davis, Donald R. and Jonathan I. Dingel, "A Spatial Knowledge Economy," NBER Working Papers 18188, National Bureau of Economic Research, Inc June 2012.

_ **and** _ , "The Comparative Advantage of Cities," NBER Working Papers 20602, National Bureau of Economic Research, Inc October 2014.

De la Roca, Jorge and Diego Puga, "Learning by Working in Big Cities," *Review of Economic Studies*, January 2017, *84* (1), 106–142.

_ **, Gianmarco Ottaviano, and Diego Puga**, "City of Dreams," *CEP Discussion Papers dp1305, Center for Economic Performance, LSE*, 2014.

Degroot, Morris H., "Reaching a Consensus," *Journal of the American Statistical Association*, 1974, *69* (345), 118–121.

Delavande, Adeline, Xavier Giné, and David McKenzie, "Measuring subjective expectations in developing countries: A critical review and new evidence," *Journal of Development Economics*, 2011, *94* (2), 151 – 163.

Desmet, Klaus, Dáavid Kristián Nagy, and Esteban Rossi-Hansberg, "The Geography of Development," *Journal Political Economy*, 2018, *126* (3), 903–983.

Diamond, Rebecca, "The Determinants and Welfare Implications of US Workers' Diverging Location Choices by Skill: 1980-2000," *American Economic Review*, 2016, *106* (3), 479–524.

_ , "The Determinants and Welfare Implications of US Workers' Diverging Location Choices by Skill: 1980-2000," *American Economic Review*, March 2016, *106* (3), 479–524.

_ **, Tim McQuade, and Franklin Qian**, "Who Benefits from Rent Control? The Equilibrium Consequences of San Francisco's Rent Control Expansion," *Working Paper*, 2019.

Dickstein, Michael and Eduardo Morales, "Estimating Dynamic Export Decisions Using Moment Inequalities," *mimeo*, April 2013.

_ **and** _ , "What Do Exporters Know?," *The Quarterly Journal of Economics*, 2018, *133* (4), 1753–1801.

Dix-Carneiro, Rafael, "Trade Liberalization and Labor Market Dynamics," *Econometrica*, 2014, *82* (3), 825–885.

_ **and Brian K. Kovak**, "Trade Liberalization and Regional Dynamics," *American Economic Review*, 2017, *107* (10), 2908–2946.

_ **and** _ , "Margins of Labor Market Adjustment to Trade," *Journal of International Economics*, 2019, *117*, 125–142.

Duranton, Gilles and Diego Puga, "Nursery Cities: Urban Diversity, Process Innovation, and the Life Cycle of Products," *American Economic Review*, December 2001, *91* (5), 1454–1477.

Dustmann, Christian, Albrecht Glitz, Uta SchÃűnberg, and Herbert BrÃijcker, "Referral-based Job Search Networks," *The Review of Economic Studies*, 10 2015, *83* (2), 514–546.

Eaton, Jonathan and Samuel Kortum, "Technology, Geography, and Trade," *Econometrica*, 2002, *70* (5), 1741–1779.

Eckert, Fabian, Mads Hejlesen, and Conor Walsh, "The Return to Big City Experience: Evidence from Danish Refugees," 2019.

Eeckhout, Jan, Roberto Pinheiro, and Kurt Schmidheiny, "Spatial Sorting," *Journal of Political Economy*, 2014, *122* (3), 554–620.

Fajgelbaum, Pablo, Eduardo Morales, Juan Carlos Suárez Serrado, and Owen Zidar, "State Taxes and Spatial Misallocation," *Review of Economic Studies*, 2019, *86*, 333–376.

Fan, Jingting, "Internal Geography, Labor Mobility, and the Distributional Impacts of Trade," *American Economic Journal: Macroeconomics*, 2019, *forthcoming.*

Farré, Lidia and Francesco Fasani, "Media Exposure and Internal Migration: Evidence from Indonesia," *Journal of Development Economics*, 2014, *102*, 48–61.

Feyrer, James, Erin T. Mansur, and Bruce Sacerdote, "Geographic Dispersion of Economic Shocks: Evidence from the Fracking Revolution," *American Economic Review*, 2017, *107* (4), 1313–1334.

Fosgerau, Mogens, Emerson Melo, Andre de Palma, and Matthew Shum, "Discrete Choice and Rational Inattention: A General Equivalence Result," *Technical Report*, May 2019.

Fujiwara, Thomas, Eduardo Morales, and Charly Porcher, "A Revealed-Preference Approach to Measuring Information Frictions in Migration Decisions," *Working Paper*, June 2020.

Gaubert, Cecile, "Firm Sorting and Agglomearation," NBER Working Papers 24478, National Bureau of Economic Research April 2018.

Giannone, Elisa, "Skill-Biased Technical Change and Regional Convergence," 2018.

Glitz, Albrecht and Rune Vejlin, "Learning Through Coworker Referrals," *Working Paper*, 2019.

Goerlich, Francisco Jose and Isidro Cantarino, "A Population Density Grid for Spain," *International Journal of Geographical Information Science*, December 2013, *27* (12), 2247–2263.

Goyal, Aparajita, "Information, Direct Access to Farmers, and Rural Market Performance in Central India," *American Economic Journal: Applied Economics*, 2010, *2* (3), 22–45.

Granovetter, Mark S., "The Strength of Weak Ties," *American Journal of Sociology*, 1973, *78* (6), 1360–1380.

Greenstone, Michael, Richard A. Hornbeck, and Enrico Moretti, "Identifying Agglomeration Spillovers: Evidence from Winners and Losers of Large Plant Openings," *Journal of Political Economy*, 2010, *118* (3), 536–598.

Grogger, Jeffrey and Gordon H. Hanson, "Income Maximization and the Selection and Sorting of International Migrants," *Journal of Development Economics*, 2011, *95* (1), 42–57.

Guiteras, Raymond, James Levinsohn, and Ahmed Mushfiq Mobarak, "Encouraging sanitation investment in the developing world: A cluster-randomized trial," *Science*, 2015, *348* (6237), 903–906.

Hastings, Justine, Christopher A Neilson, and Seth D Zimmerman, "The Effects of Earnings Disclosure on College Enrollment Decisions," Working Paper 21300, National Bureau of Economic Research June 2015.

Humlum, Anders, "Robot Adoption and Labor Market Dynamics," *Working Paper*, 2020.

Jensen, Robert, "The Digital Provide: Information (Technology), Market Performance, and Welfare in the South Indian Fisheries Sector*," *The Quarterly Journal of Economics*, 08 2007, *122* (3), 879–924.

_ , "The (Perceived) Returns to Education and the Demand for Schooling," *Quarterly Journal of Economics*, 2010, *125* (2), 515–548.

_ **and Nolan H Miller**, "Keepin' 'em Down on the Farm: Migration and Strategic Investment in Children's Schooling," Working Paper 23122, National Bureau of Economic Research February 2017.

Joo, Joonwhi, "Buying a Larger Package with Quantity Surcharge: Information Friction or Preference Heterogeneity," *Working Paper*, 2017.

Kaplan, Greg and Sam Schulhofer-Wohl, "Understanding the Long-run Decline in Interstate Migration," *International Economic Review*, 2017, *58* (1).

Kendall, Chad, Tommaso Nannicini, and Francesco Trebbi, "How Do Voters Respond to Information? Evidence from a Randomized Campaign," *American Economic Review*, January 2015, *105* (1), 322–53.

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

Koudijs, Peter, "Those Who Know Most: Insider Trading in Eighteenth-Century Amsterdam," *Journal of Political Economy*, 2015, *123* (6), 1356–1409.

_ , "The Boats That Did Not Sail: Asset Price Volatility in a Natural Experiment," *The Journal of Finance*, 2016, *71* (3), 1185–1226.

Kovak, Brian K, "Regional Effects of Trade Reform: What is the Correct Measure of Liberalization?," *American Economic Review*, 2013, *103* (5), 1960–76.

Kramarz, Francis and Oskar NordstrÃűm Skans, "When Strong Ties are Strong: Networks and Youth Labour Market Entry," *The Review of Economic Studies*, 01 2014, *81* (3), 1164–1200.

Levy, Gilat and Ronny Razin, "Information diffusion in networks with the Bayesian Peer Influence heuristic," *Games and Economic Behavior*, 2018, *109*, 262 – 270.

Manning, Alan, "The Plant Size-Place Effect: Agglomeration and Monopsony in Labour Markets," *Journal of Economic Geography*, September 2010, *10* (5), 717–744.

Manski, Charles F., "Nonparametric Estimation of Expectations in the Analysis of Discrete Choice Under Uncertainty," in William Barnett, James Powell, and George Tauchen, eds., *Nonparametric and Semiparametric Methods in Econometrics and Statistics*, Cambridge: Cambridge University Press, 1991.

_ , "Adolescent econometricians: How do youth infer the returns to schooling?," in Charles Clotfelter and Michael Rothschild, eds., *Studies of Supply and Demand in Higher Education*, Chicago: University of Chicago Press, 1993, pp. 43–60.

_ , "Measuring Expectations," *Econometrica*, 2004, *5* (9), 1329–1376.

_ , *Identification for Prediction and Decision*, Cambridge: Harvard University Press, 2007.

Marques, Filipe J., Carlos A. Coelho, and Miguel de Carvalho, "On the distribution of linear combinations of independent Gumbel random variables," *Statistics and Computing*, 5 2015, *25* (3), 683–701. WOS:000352705000013.

Matejka, Filip and Alisdair McKay, "Rational Inattention to Discrete Choices: A New Foundation for the Multinomial Logit Model," *American Economic Review*, January 2015, *105* (1), 272–98.

Matsuyama, Kiminori and Philip Ushchev, "Beyond CES: Three Alternative Classes of Flexible Homothetic Demand Systems," *mimeo*, July 2017.

McCauley, Jeremy, "The Role of Information in Explaining the Lack of Welfare-Induced Migration," *mimeo*, January 2019.

McKenzie, David, John Gibson, and Steven Stillman, "A land of milk and honey with streets paved with gold: Do emigrants have over-optimistic expectations about incomes abroad?," *Journal of Development Economics*, 2013, *102*, 116 – 127. Migration and Development.

Mellow, Wesley, "Employer Size and Wages," *The Review of Economics and Statistics*, August 1982, *64* (3), 495–501.

Mian, Atif and Amir Sufi, "What Explains the 2007-2009 Drop in Employment?," *Econometrica*, 2014, *82* (6), 2197–2223.

Miller, Grant and A. Mushfiq Mobarak, "Learning About New Technologies Through Social Networks: Experimental Evidence on Nontraditional Stoves in Bangladesh," *Marketing Science*, 2015, *34* (4), 480–499.

Molavi, Pooya, Alireza Tahbaz-Salehi, and Ali Jadbabaie, "A Theory of Non-Bayesian Social Learning," *Econometrica*, 2018, *86* (2), 445–490.

Monte, Ferdinando, Stephen J. Redding, and Esteban Rossi-Hansberg, "Commuting, Migration and Local Employment Elasticities," *American Economic Review*, 2018, *108* (12), 3855–3890.

Morales, Eduardo, Gloria Sheu, and Andrés Zahler, "Extended Gravity," *mimeo*, June 2017.

Moretti, Enrico, "Local Labor Markets," in Orley Ashenfelter and David Card, eds., *Handbook of Labor Economics*, Amsterdam: Elsevier North-Holland, 2011, pp. 1238–1303.

Morten, Melanie and Jaqueline Oliveira, "The Effects of Roads on Trade and Migration: Evidence from a Planned Capital City," *mimeo*, 2018.

Mueller, Holger M., Paige P. Ouimet, and Elena Simintzi, "Wage Inequality and Firm Growth," *American Economic Review*, May 2017, *107* (5), 379–83.

Nadarajah, Saralees, "Exact distribution of the linear combination of p Gumbel random variables," *International Journal of Computer Mathematics*, 2008, *85* (9), 1355–1362.

Nakamura, Eli, Jósef Sigurdsson, and Jón Steinsson, "The Gift of Moving: Intergenerational Consequences of a Mobility Shock," *Working Paper*, 2019.

Nakamura, Emi and Jón Steinsson, "Fiscal Stimulus in a Monetary Union: Evidence from US Regions," *American Economic Review*, 2014, *104* (3), 753–792.

Perihan, Andrea Weber Saygin and Michele Weynandt, "Coworkers, Networks, and Job Search Outcomes," *ILR Review*, 2014.

Pierce, Justin R. and Peter K. Schott, "Trade Liberalization and Mortality: Evidence from US Counties," *American Economic Review: Insights*, 2018, *forthcoming.*

Porcher, Charly, "Migration with Costly Information," *Working Paper*, 2020.

Powell, Warren, *Approximate Dynamic Programming: Solving the Curses of Dimensionality*, second ed., New York: John Wiley and Sons, 2011.

Redding, Stephen J., "Goods Trade, Factor Mobility and Welfare," *Journal of International Economics*, 2016, *101*, 148–167.

Restrepo-Echavarria, Paulina and Brian Reinbold, "Measuring Labor Share in Developing Countries," *Federal Reserve Bank of St. Louis The Regional Economist, First Quarter*, 2018, *26* (1), 31–58.

Roy, A. D., "Some Thoughts on the Distribution of Earnings," *Oxford Economic Papers, New Series*, 1951, *3* (2), 135–146.

Schmutte, Ian M., "Job Referral Networks and the Determination of Earnings in Local Labor Markets," *Journal of Labor Economics*, 2015, *33* (1), 1–32.

Scott, Paul, "Dynamic Discrete Choice Estimation of Agricultural Land Use," *Working Paper*, 2013.

Shrestha, Maheshwor, "Get Rich or Die Tryin': Perceived Earnings, Perceived Mortality Rate and the Value of a Statistical Life of Potential Work-Migrants from Nepal," *World Bank Policy Research Working Paper*, 2017, *7945*.

Sims, Christopher A., "Implications of rational inattention," *Journal of Monetary Economics*, 2003, *50* (3), 665 – 690.

Sjaastad, Larry A., "The Costs and Returns of Human Migration," *Journal of Political Economy*, 1962, *70* (5), 80–93.

Small, Kenneth A. and Harvey S. Rosen, "Applied Welfare Economics with Discrete Choice Models," *Econometrica*, 1981, *49* (1), 105–130.

Steiner, Jakub, Colin Stewart, and Filip MatÄŻjka, "Rational Inattention Dynamics: Inertia and Delay in Decision-Making," *Econometrica*, 2017, *85* (2), 521–553.

Steinwender, Claudia, "Real Effects of Information Frictions: When the States and the Kingdom Became United," *American Economic Review*, March 2018, *108* (3), 657–96.

Suárez Serrato, Juan Carlos and Owen Zidar, "Who Benefits from State Corporate Tax Cuts? A Local Labor Markets Approach with Heterogeneous Firms," *American Economic Review*, 2016, *106* (9), 2582–2624.

Tian, Lin, "Division of Labor and Productivity Advantage of Cities: Theory and Evidence from Brazil," *Working Paper*, 2019.

Tombe, Trevor and Xiaodong Zhu, "Trade, Migration and Productivity: A Quantitative Analysis of China," *American Economic Review*, 2019, *forthcoming*.

Topalova, Petia, "Factor Immobility and Regional Impacts of Trade Liberalization: Evidence on Poverty from India," *American Economic Journal: Applied Economics*, 2010, *2* (4), 1–41.

Traiberman, Sharon, "Occupations and Import Competition: Evidence from Denmark," *mimeo*, March 2019.

van der Klauuw, Wilbert, "On the Use of Expectations Data in Estimating Structural Dynamic Choice Models," *Journal of Labor Economics*, 2012, *30* (3), 554–571.

Wilson, Riley, "Moving to Jobs: The Role of Information in Migration Decisions," *mimeo*, December 2018.

Yagan, Danny, "Employment Hysteresis from the Great Recession," *Journal of Political Economy*, 2019, *forthcoming*.

Zimmerman, Seth D., "Elite Colleges and Upward Mobility to Top Jobs and Top Incomes," *American Economic Review*, January 2019, *109* (1), 1–47.