Migration with Costly Information

Charly Porcher*

This version: April 2022
First version: October 2019

Abstract

Information is critical for migration decisions. Yet, depending on where individuals reside and who they interact with, they may face different costs of accessing information about employment opportunities. How does this imperfect and heterogeneous information structure affect the spatial allocation of economic activity and welfare? To address this question, I develop a quantitative dynamic model of migration with costly information acquisition and local information sharing. Rationally inattentive agents optimally acquire more information about nearby locations and learn about payoffs in other locations from the migrants around them. I apply this model to internal migration in Brazil and estimate it using migration flows between regions. To illustrate its quantitative implications, I evaluate the counterfactual effects of the roll-out of broadband internet in Brazil. By allowing workers to make better mobility choices, expanding internet access increases average welfare by 1.6%, reduces migration flows by 1.2% and reduces the cross-sectional dispersion in earnings by 4%.

*McDonough School of Business, Georgetown University. Contact: cp1105@georgetown.edu. I am extremely grateful to my advisors Esteban Rossi-Hansberg, Eduardo Morales and Jakub Kastl for their continuous support and guidance. I also thank Adrien Bilal, Damien Capelle, Thomas Fujiwara, Gene Grossman, Federico Huneeus, Anders Humlum, Oleg Itskhoki, Adam Kapor, Nobuhiro Kiyotaki, Adrien Matray, Thierry Mayer, Mathilde Le Moigne, Ezra Oberfield, Fernando Parro, Leah Platt Boustan, Steve Redding, Richard Rogerson, Chris Sims, as well as numerous seminar participants for helpful discussions and comments. This research benefited from financial support from the International Economics Section (IES) at Princeton University. Its support is gratefully acknowledged.

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. Are people missing many opportunities because of their lack of information? Is there inequality in access to information and does it matter for inequality in economic outcomes? How did the large improvements in access to information over the past couple of decades affect the way people react to regional shocks? What is the scope for policies to affect the structure of migrants' information and to induce them to take better advantage of migration opportunities?

In this paper, I propose answers to these questions. I start by arguing that we need to consider what migrants know about opportunities elsewhere not simply as a primitive, but as an equilibrium outcome. Indeed, some locations offer better infrastructure for collecting information than others. Migrants can also share information about past migrations with their neighbors. Since migration partly shapes local networks and relocates workers, the structure of information itself may be altered by migration patterns.

I then proceed with four contributions. First, I develop a theory of spatial equilibrium with migration in the presence of information frictions, in which the equilibrium structure of workers' information is determined jointly with migration patterns. 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 the 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, location-specific productivity shocks that are unobserved to agents alter the spatial distribution of earnings. Agents are rationally inattentive and can acquire information about these shocks at some cost, which can vary by location. They 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 main 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 in particular. By enhancing flows to closer destinations at the expense of more remote ones, these biases 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 payoffs that are unobserved by agents, such as time-varying shocks to local earnings. Recovering both the elasticity of migration with respect to time-varying productivity shocks and the elasticity with respect to time-invariant components of earnings 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. Each period, the state variables consist of the vector of productivity shocks in each location, the matrix of bilateral migration flows between pairs of locations, 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 beliefs can be described as a function of migration flows only, reducing the effective state space to the productivity vector and migration flows matrix. 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.

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 the 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 the 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

[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.

acquire information about other regions. These 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 close to 3% of earnings.[3] Hence, collecting information to decide where to move appears almost as costly as moving itself. Importantly, the estimated migration costs would be 40% larger if I ignored information frictions by imposing that all information costs were equal to zero.[4] 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, I find empirically that 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. The migration flows predicted by the model also possess these features, in comparable magnitudes. The heterogeneous responsiveness of mobility decisions in the simulated migration patterns arises because agents' beliefs are more tightly correlated with the true payoffs in regions with which they have more interactions. Second, I reproduce empirical findings obtained in Fujiwara, Morales and Porcher (2019) and show 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 also arises because of 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 stochastic 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 stochastic steady state with complete information, gross migration flows are more concentrated towards high payoffs regions and are 4.1% lower. Second, I evaluate the counterfactual effect of the roll-out of broadband internet in Brazil during the early 2010s.[5] 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

[3] I estimate the average migration cost to be about 75% of annual earnings. However, the median worker only migrates once in a lifetime. With an annual discount rate of 0.96, paying 75% of earnings in one year is comparable to paying 3% of earnings every year.

[4] 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."

[5] I abstract from the direct effects that broadband internet access may have had on local productivity and focus only on its role in improving migrants' information. Broadband internet still affects aggregate productivity, but only through the reallocation of migration flows.

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 improved 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, despite benefitting from a better access to internet. These regions suffer for two reasons. First, their internet access increased less than in most other regions. As a result, workers in other locations are relatively more informed and can move faster to take advantage of any local opportunity, making it less likely that workers in the remote regions can enjoy the high wages in booming regions. Second, fewer well-informed workers move to these relatively unattractive remote 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, 2013; Wilson, 2018), or by directly providing information about average earnings in other locations in a randomized experiment (Baseler, 2019; Bryan et al., 2014).[6] Second, a recent body of work documents that 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 (2019) use a revealed preference approach to infer the composition of migrants' information consistent with a model of migration allowing for a rich set of unobserved 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.,

[6] 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 this information, or the information made available was not useful or credible. Consistent with the first interpretation, Fujiwara, Morales and Porcher (2019) 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.

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), Matějka and McKay (2015) showed how static rational inattention problems lead to a multinomial logit decision rule.[7] 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 (2018) provided a microfoundation for the model from Eaton and Kortum (2002).[8] In the first application to migration, Bertoli et al. (2019) show that international migration flows feature heterogeneous elasticities, consistent with rational inattention.[9] 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.[10] In their analysis of internal migration patterns in the United States, Kennan 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.[11] Such large migration costs have been shown to have important implications for aggregate labor productivity and welfare.[12] 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

[7]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.

[8]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.

[9]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.

[10]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).

[11]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.

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

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, I present my model. Section 3 describes the data I will be using for the estimation. In Section 4, I carry out the estimation and discuss the results. Section 5 confronts some of the model's predictions to the data. Section 6 presents the counterfactual exercises.

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.

2.1 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.[13] Let $\Gamma(\theta_t|\theta^{t-1})$ 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).

[13] As discussed in Footnote 10, 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 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.

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

$$u_{ijkt} = u(c_{kt}) + B_k - \kappa_{jk} + \nu\varepsilon_{ikt}, \tag{1}$$

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

$$y_{kt} = \exp(A_k + \theta_{kt})L_{kt}^{1-\alpha}, \qquad \alpha > 0, \tag{2}$$

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

$$\theta_{jt} = \rho\theta_{jt-1} + \xi_{jt}, \qquad \xi_{jt} \sim \mathcal{N}\left(0, \sigma_\xi^2\right). \tag{3}$$

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:

$$\log w_{kt} = A_k + \theta_{kt} - \alpha \log L_{kt} + \log\left(1-\alpha\right). \tag{4}$$

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 $w_{kt}/(1-\alpha)$.[14] Moreover, the assumption that the economy features free trade

[14] 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 $\sum_k \tau_t w_{kt} L_{kt} =$

in goods between locations implies that local consumption prices are equalized across locations and do not affect migration decisions. As a result, indirect utility net of information costs can be expressed as

$$u_{ijkt} = A_k + \theta_{kt} - \alpha \log L_{kt} + B_k - \kappa_{jk} + \nu\varepsilon_{ikt}. \tag{5}$$

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 variable that is imperfectly observed by agents.[15] 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. 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:

$$\pi_{it}(\theta) = \Pr(\theta_t = \theta | i, t), \qquad \forall \theta \in \mathbb{R}^J. \tag{6}$$

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 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 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 2.3, 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 2.4. 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}.

$\sum_k \alpha \exp\left(A_k + \theta_{kt}\right) L_{kt}^{1-\alpha}$, leading to $\tau = \alpha/(1-\alpha)$ after substituting wages by their expression in (4).

[15] 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.

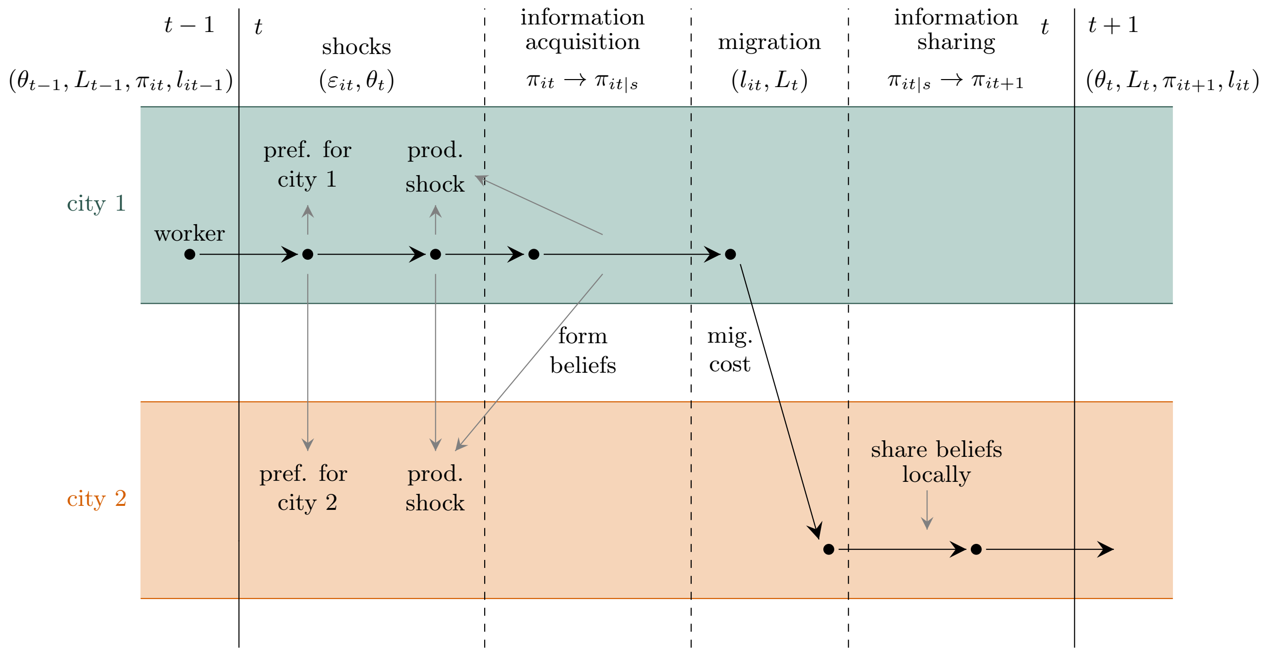


Figure 1: Sequence of actions during any period t. Example of a worker moving from city 1 to city 2.

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 Matějka 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 them 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 their signal structure for the period, nature draws a signal realization s from $f(\cdot)$. Given the signal, the agent updates their prior, resulting in the posterior belief $\pi_{it|s}$ after applying Bayes' rule:

$$\pi_{it|s}(\theta_t) = \frac{f(s|\theta_t, \omega_{it})\pi_{it}(\theta_t)}{f(s|\omega_{it})}. \tag{7}$$

To describe the costs of acquiring information, I follow the rational inattention literature and rely on the entropy of beliefs to measure their uncertainty.[16] 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

$$H(p) \equiv \tilde{H}(X) = -\int_{x \in S} p(x) \log p(x) dx. \tag{8}$$

Entropy is a measure of uncertainty about X.[17] 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.[18] I define the cost of designing a signal strategy f for agent with state ω_{it} to be

$$I\left(f|\omega_{it}\right) = \lambda_{l_{it-1}}\left(H\left(\pi_{it}\right) - \mathbb{E}_s\left[H\left(\pi_{it|s}\right)\right]\right), \tag{9}$$

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

$$\max_{f,\sigma} \mathbb{E}\left[\sum_{t=1}^{\infty} \delta^t \left(u_{l_{it-1}l_{it}}(\theta_t, L_t, \varepsilon_{it}) - I(f_t|\omega_{it})\right)\right] \quad s.t. \quad l_{it} = \sigma_t(s_{it}, \omega_{it}), \tag{10}$$

where l_{it} is the location optimally chosen at t, the flow of utility is defined in (5), the information cost is defined in (9), and the expectation is taken 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 A.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

[16] 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.

[17] 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.

[18] 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.

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 shock θ_{kt}.[19]

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

$$p_{jkt}(s, \theta_t, \omega_{it}) = f_t(s|\theta_t, \omega_{it})\mathbb{1}\{\sigma_t(s, \omega_{it}) = k\}\mathbb{1}\{l_{it-1} = j\}. \tag{11}$$

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

$$L_{jkt} = L_{jt-1}\bar{p}_{jkt}(\theta_t, L_{t-1}), \tag{12}$$

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.[20]

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] 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).

[20] 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 (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 (12). In Appendix A.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.

2.5 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.[21] 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

$$\log \bar{\pi}_{kt}(\theta_t) = C_{kt} + \sum_j \sum_s L_{jkt|s} \log \pi_{jt|s}(\theta_t), \tag{13}$$

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.[22] 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.[23] In Appendix A.1.1, I follow Molavi et al.

[21] 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.

[22] 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.

[23] See Appendix A.1.1.

(2018) and show that the log-linear rule (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 beliefs at the end of t,

$$\pi_{jt+1}(\theta_{t+1}) = \int_\theta \bar{\pi}_{jt}(\theta)\gamma(\theta_{t+1}|\theta)d\theta, \tag{14}$$

where $\gamma(\cdot|\theta)$ is the pdf of a normal distribution with mean $\rho\theta$ and variance σ_ξ^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 states that are observed by agents $\omega_{it} = (L_{t-1}, \pi_{l_{it-1}t}, l_{it-1}, \varepsilon_{it})$.

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 (10).*
- ***Mobility**: Population evolves as in (12).*
- ***Beliefs**: Posterior beliefs are derived from priors according to Bayes' rule (7), and are shared locally according (13), leading to next-period prior beliefs about future productivity in (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.[24]

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 (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 (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 (10) generates a choice rule p that solves*

$$\max_{p} \mathbb{E}\left[\sum_{t=1}^{\infty} \delta^t \left(u_{l_{t-1}l_t}(\theta_t, \omega_{it}) - I(\omega_{it})\right)\right], \tag{15}$$

where the information cost defined in (16) is expressed as a function of the prior and posterior beliefs $\pi_{jkt} \equiv \pi_{jt|s_t=k}$:

$$I(\omega_{it}) = \lambda_{l_{it-1}} \left(H(\pi_{jt}) - \sum_k q_{jkt}(\omega_{it}) H(\pi_{jkt})\right) \qquad \forall j \tag{16}$$

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 (12) – (14).

Proof. See Appendix A.1.2. □

Accordingly, I will dispense with the signals s_t, replacing them with actions l_t, and will refer to any rule p solving (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 A.1.3. □

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.

[24] 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.

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* (15)*, while population and beliefs follow the laws of motion* (12) *–* (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 A.1.4, 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).[25]

I am now ready to characterize the solution to the stochastic steady-state. As was shown by Matějka 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 (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*

$$p_{jk}(\theta_t, L_{t-1}, \varepsilon_{it}) = \frac{q_{jk}(L_{t-1}, \varepsilon_{it}) \exp\left(u_{jk}(\theta_t, L_{t-1}, \varepsilon_{it}) + \delta \bar{V}_k(\theta_t, L_{t-1})\right)^{1/\lambda_j}}{\sum_l q_{jl}(L_{t-1}, \varepsilon_{it}) \exp\left(u_{jl}(\theta_t, L_{t-1}, \varepsilon_{it}) + \delta \bar{V}_l(\theta_t, L_{t-1})\right)^{1/\lambda_j}} \tag{17}$$

[25] 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}.

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*

$$V_j(\theta_t, L_{t-1}, \varepsilon_{it}) = \lambda_j \log\left(\sum_l q_{jl}(L_{t-1}, \varepsilon_{it}) \exp\left(u_{jl}(\theta_t, L_{t-1}, \varepsilon_{it}) + \delta\bar{V}_l(\theta_t, L_{t-1})\right)^{1/\lambda_j}\right), \tag{18}$$

and population and beliefs follow the laws of motion (12) *–* (14)*.*

Proof. See Appendix A.1.4. □

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 they obtain. 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 (17) as $\lambda_j \to 0$. When information is costly and $\lambda_j > 0$, the expression (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 (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 (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 locations. 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.[26] 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.[27]

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}):

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

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*

$$\bar{p}_{jk}(\theta_t, L_{t-1}) = \frac{\exp\left(\eta_j \chi_{jk}(\theta_t, L_{t-1}) + \bar{v}_{jk}(\theta_t, L_{t-1})\right)^{\frac{1}{\phi_j}}}{\sum_l \exp\left(\eta_j \chi_{jl}(\theta_t, L_{t-1}) + \bar{v}_{jl}(\theta_t, L_{t-1})\right)^{\frac{1}{\phi_j}}}, \tag{19}$$

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

$$V_j(\theta_t, L_{t-1}) = \phi_j \log\left(\sum_l \exp\left(\eta_j \chi_{jl}(\theta_t, L_{t-1}) + \bar{v}_{jl}(\theta_t, L_{t-1})\right)^{\frac{1}{\phi_j}}\right), \tag{20}$$

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

[26] 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 (2018) imposed that productivity shocks follow a particular type of conjugate EV1, the Cardell distribution, in their model of international trade.

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

preference-based migration model:

$$\bar{p}_{jk}(\theta_t, L_{t-1}) \rightarrow_{\lambda_j \rightarrow 0} \frac{\exp\left(\bar{v}_{jk}(\theta_t, L_{t-1})\right)^{1/\nu}}{\sum_l \exp\left(\bar{v}_{jl}(\theta_t, L_{t-1})\right)^{1/\nu}}. \tag{21}$$

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

$$\bar{p}_{jk}(\theta_t, L_{t-1}) = \frac{\exp\left(\rho_j \chi_{jk}(\theta_t, L_{t-1}) + \bar{v}_{jk}(\theta_t, L_{t-1})\right)^{\frac{1}{\psi_j}}}{\sum_l \exp\left(\rho_j \chi_{jl}(\theta_t, L_{t-1}) + \bar{v}_{jl}(\theta_t, L_{t-1})\right)^{\frac{1}{\psi_j}}}, \tag{22}$$

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

Proof. See Appendix A.1.5. □

The role of information frictions λ_j in altering the responsiveness of migration to local shocks is apparent in (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, $\eta_j \rightarrow 1$ 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 (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 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 (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 (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 (17). As I will discuss in Section 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 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.

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.

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.[28] 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 spell, the average monthly wage, the number of hours stipulated in the contract, and certain characteristics of the plant where the worker is employed.[29] Specifically, I use information on the micro and mesoregion in which the plant is located.[30]

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

[28] In Brazil, informal and self-employed workers constitute a large fraction of the labor force, peaking at 45% in 2000 (Bosch 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.

[29] In the analysis of migration patterns at the individual level described in section 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.

[30] 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.

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 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%.

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 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.

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.[31] 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

[31] 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.

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 (4) and the evolution of productivity (3), the adjusted wages can be expressed as

$$\tilde{w}_{kt} \equiv \log w_{kt} + \alpha \log L_{kt} - \log(1-\alpha) = A_k + \theta_{kt}. \tag{23}$$

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 $(1-\rho^2)\mathrm{Var}(\tilde{w}_{kt})$. Finally, I recover the productivity shocks θ_{kt} implied from the observed $\tilde{w}_{kt}$ and estimated productivities A_k as the residuals from (23).[32] As reported in Table 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^2_ξ is 0.33 in the first period and 0.38 in the second period.

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.[33] 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 (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 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 (5) yields the following expression:

$$\log \frac{\bar{p}_{jkt}}{\bar{p}_{jjt}} = \frac{\eta_j}{\phi_j} \left(\chi_{jkt} - \chi_{jjt}\right) + \frac{1}{\phi_j} \left(\log \frac{w_{kt}}{w_{jt}} + D_{jk} + \delta \left(V_{kt+1} - V_{jt+1}\right) \right) + e_{jkt}, \tag{24}$$

[32] 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.

[33] 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.

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 $e_{jkt} = (\zeta_{jjt} - \zeta_{jkt})/\phi_j$ is composed of expectational errors about future values $\zeta_{jkt} = \delta (V_{kt+1} - E_{j,t}V_{kt+1})$. Equation (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 (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 $\vartheta = (\nu, \lambda_j, B_j, \kappa_{jk})$. Intuitively, 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.[34]

The estimation algorithm I employ relies on a standard fixed point iteration on the OLS equation (24). A satisfactory set of parameters ϑ should deliver beliefs and values such that the estimates $\hat{\beta}$ from (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.[35] 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 (24) by OLS.[36] 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 (24). I describe the iterative algorithm in more detail in Appendix C.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 1. I describe the information costs in Section 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 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

[34] 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 C.1 for more details.

[35] To obtain reasonable initial guesses, I first run (24) omitting the mean beliefs and future values.

[36] 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 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.

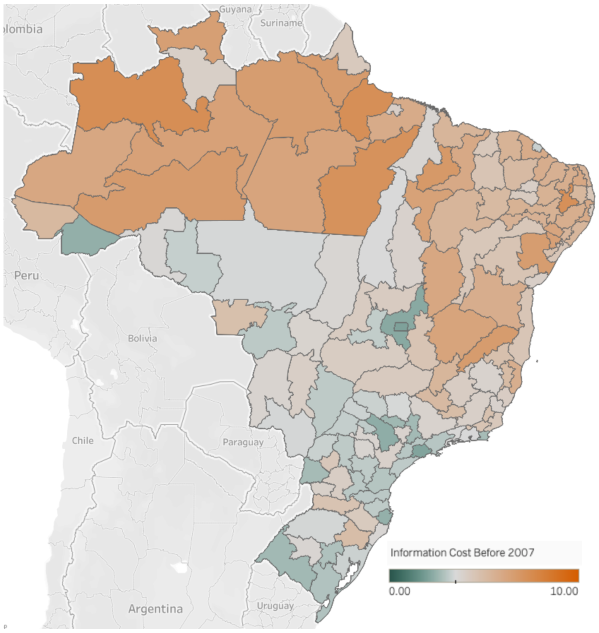
obtain that the standard deviation is 0.83 in the first period and 0.91 in the second period. The average migration costs are 0.56 in the first period and 0.47 in the second period. The decline in migration costs between the two periods is consistent with the increase in migration flows between the two periods.[37] Overall, parameter values are fairly similar over the two periods.

4.3 Information Costs and Internet Access

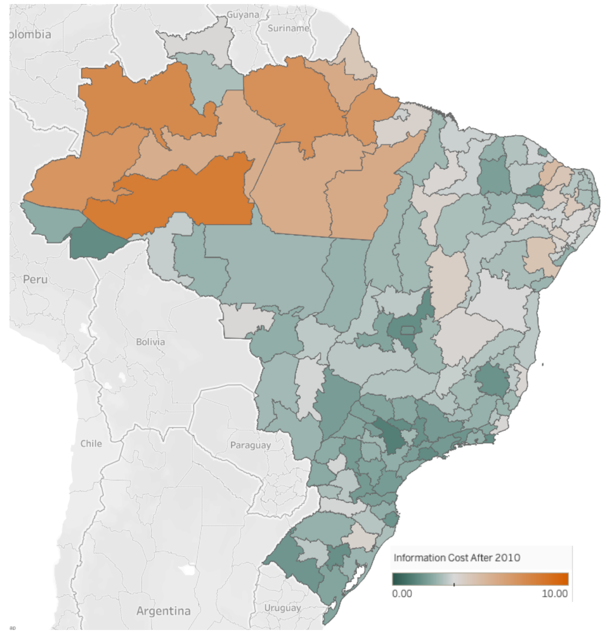
Information costs are on average equal to 3.11 in the first period and 2.23 in the second period. The standard deviation in the information costs is 1.43 in the first period and 1.32 in the second period. Figure 2a and Figure 2b display the estimated information costs λ_j for each of the 137 mesoregions for the first and second periods. The information costs appear strongly correlated with the economic development 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 2a and Figure 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 propose a simple parameterization of information

[37]See Appendix B.2 for a description of the yearly migration rates.



(a) Information Costs 2000-2007



(b) Information Costs 2008-2014

costs:

$$\lambda_{jt} = \ell_1 int_{jt} + \ell_2 \log w_{jt} + \ell_3 \log popdens_{jt} + \varsigma_j + u_{jt}, \tag{25}$$

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 (25) in first differences by OLS and report the results in the first column of Table 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 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 (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

Table 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.

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.[38] 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 C.2. I report the results from the instrumental variable regression in column 2 of Table 2. The magnitude of the coefficient of internet connections per resident is slightly lower, but remains large and significant at the 5% level.

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 (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

[38]"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.

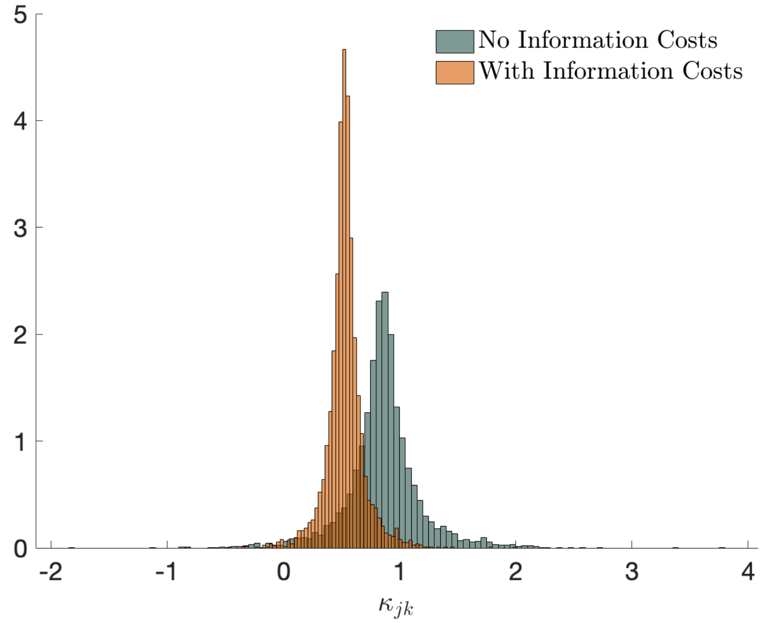


Figure 3: Estimated Migration Costs with and without Information Cost

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

$$\Lambda_{jkt} = \frac{1}{\nu}\left(\Delta \log w_{jkt} + D_{jk} + e_{jkt}\right), \tag{26}$$

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:

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

The iterative estimation algorithm described in Section 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 (26)—after projecting on origin and destination fixed effects to net out amenities—is displayed in Figure 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 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 (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 (21), Caliendo et al. (2018) estimate $\nu = 2.43$ in the context of migration between European countries during a similar period.

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:

$$\kappa_{jk} = \beta_1 \log dist_{jk} + \beta_2 \log traveltime_{jk} + \beta_3 contiguous_{jk} + e_{jk}. \tag{27}$$

Table 3 reports the results from estimating (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.

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

Table 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.

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 (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.

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 (21), there is a straightforward strategy for recovering the migration elasticity $1/\nu$, described in (26). I can however allow for some heterogeneity in the migration elasticity along some variable Z_{jkt}:

$$\Lambda_{jkt} = \beta_1 \Delta \log w_{jkt} + \beta_2 Z_{jkt} \times \Delta \log w_{jkt} + \beta_3 Z_{jkt} + D_{jk} + e_{jkt}. \tag{28}$$

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 (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 4, I report the results from estimating equation (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 columns 1, 4, and 7, I report the results from estimating equation (28) with observed migration flows and wages over the period 2008–2014. Columns 2, 5, and 8 contain the

Table 4: Heterogeneity of Migration Elasticities in the Data and in the Model

	(1) Data	(2) Model Info	(3) Model No Info	(4) Data	(5) Model Info	(6) Model No Info	(7) Data	(8) Model Info	(9) Model No Info
Income gap	0.391^a	0.342^a	0.362^a	0.398^a	0.306^a	0.319^a	0.292^a	0.241^a	0.311^a
	(0.012)	(0.007)	(0.008)	(0.016)	(0.008)	(0.009)	(0.018)	(0.008)	(0.008)
Log dist	-0.012^a	-0.009^a	0.002						
$\times$ Inc. gap	(0.004)	(0.002)	(0.002)						
Internet				0.132^a	0.101^a	0.003			
$\times$ Inc. gap				(0.015)	(0.004)	(0.004)			
Log Past Flows							0.003^a	0.006^a	0.000
$\times$ Inc. gap							(0.001)	(0.001)	(0.001)
Observations	68,152	68,152	68,152	68,152	68,152	68,152	68,152	68,152	68,152
R^2	0.179	0.515	0.613	0.157	0.513	0.614	0.168	0.582	0.634

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.

results from estimating equation (28) with migration flows and wages predicted by the model using the estimated parameters in the period 2008–2014. Once I solve for the 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. Columns 3, 6, and 9 report the results from estimating (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 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 4.4. There is, therefore, significant variation of the migration elasticity with distance.[39] In 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

[39] 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.

between remote regions.

In column 4, 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 column 5, I report the results from running the 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 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 column 6, 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 column 6, 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.

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. (2019) to study the migration response to local shocks in Brazil.

In the model with perfect information (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

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

I also define the expected time of migration between j and k over a horizon $(\underline{t}, \bar{t})$ as $\sum_{s=0}^{\bar{t}} s p^{(\underline{t},\bar{t})}_{jkt+s}$. 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 D.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:

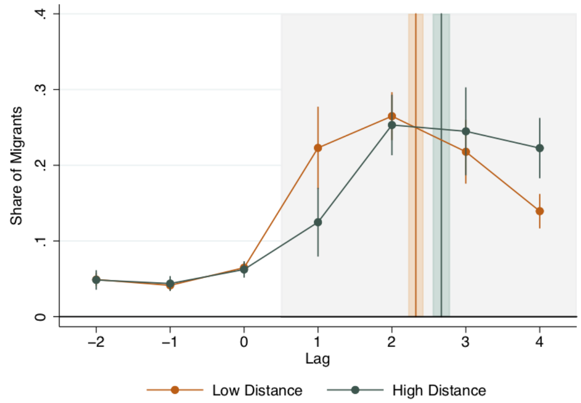
$$p^{(\underline{t},\bar{t})}_{jkt+s} = \sum_{s'=-2}^{4} \mathbb{1}\{s = s'\} \left(\gamma_{1s'} int_{jt} + \gamma_{2s'} \log dist_{jk} + \gamma_{3s'} \log p^{past}_{jk} \right) + u_{jkt+s}, \tag{29}$$

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^{past}_{jk} 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.

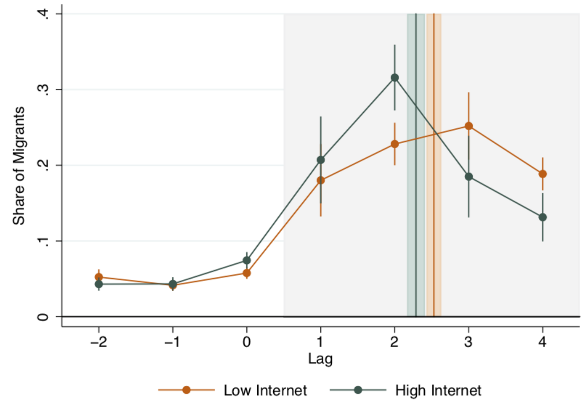
Figure D.2a and Figure D.2b display the estimates of $\{\gamma_{1t+s}, \gamma_{2t+s}\}$ that indicates how internet access, distance and past migration flows affect the timing of migration. I focus here on the largest labor demand shock in the sample, which took place in Ipojuca in the region of Recife 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 25^{th} percentile, labeled as "Low", or to its 75^{th} percentile, labeled as "High", 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 (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

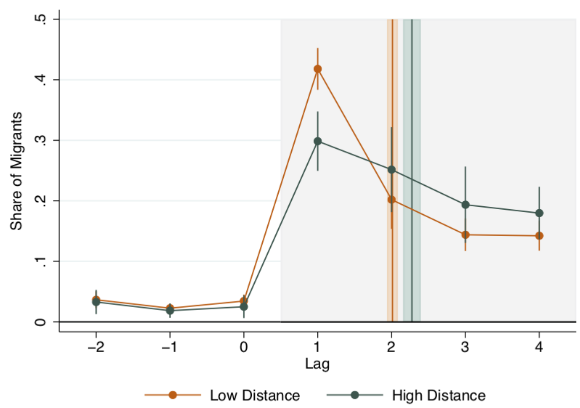
Figure 4: Delay in the Migration Response to the Local Shock in Ipojuca



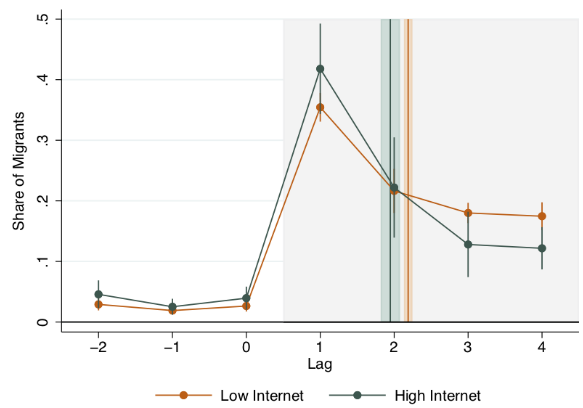
(a) Distance: Data



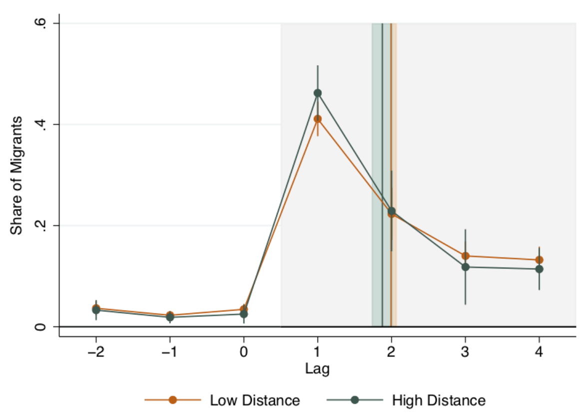
(b) Internet: Data



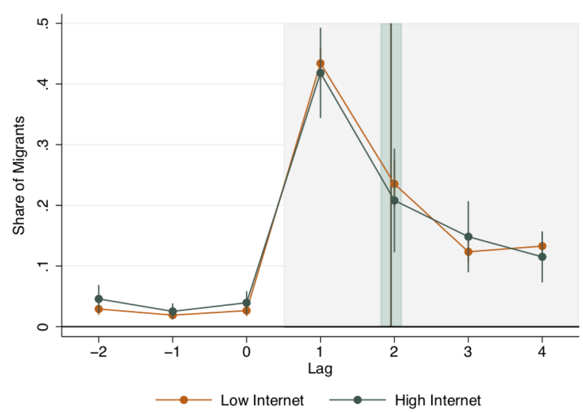
(c) Distance: Model with information frictions



(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.

the model under the observed population and recovered productivity vectors. The recovered productivity features a persistent increase starting in 2009 (see Appendix D.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 4c illustrates that the estimated coefficients on distance lead to a significant delay for regions that are farther from to Recife. Figure 4d shows that the model also predicts a significant delay from regions with lower internet access.

In the model with no information frictions, Figures 4e and 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 D.4, I show that this differential delay holds for other local shocks occurring in Brazil between 2000 and 2014.

6 Counterfactual Exercises

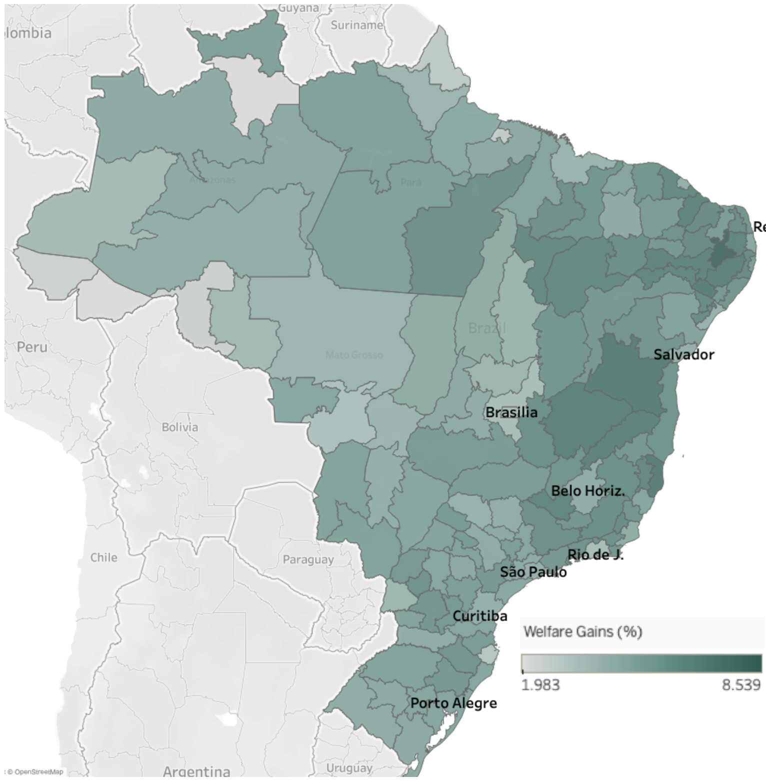
I have discussed several implications of information frictions on migration patterns. The main results of interest are that individuals feature endogenous predispositions towards moving to some regions, leading to lower estimated migration costs, and that their responsiveness to variations in payoffs 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.

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 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

Figure 5: Welfare Gains by Mesoregion after Setting All Information Costs to Zero



distribution. Denoting by $\mathcal{V}_j$ the expected value of a region,

$$\mathcal{V}_j = \mathbb{E}_{\theta_t, L_{t-1}}\left[\sum_k \bar{p}_{jk}(\theta_t, L_{t-1})\left(\bar{u}_{jk}(\theta_t, L_{t-1}) + \delta \bar{V}_k(\theta_t, L_{t-1})\right) - I_j(\pi_j(L_{t-1}))\right]. \tag{30}$$

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 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 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 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 $\bar{U}_j = (\bar{u}_{jk} + \delta\bar{V}_k)_{k=1,\ldots,J}$ 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 $\bar{p}_j = (\bar{p}_{jk})_{k=1,\ldots,J}$ 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 (30) as

$$\Delta\mathcal{V}_j = \underbrace{\mathbb{E}_{\theta,L}\left[\Delta\bar{p}_j \cdot \bar{U}_j'\right]}_{\text{better sorting}} + \underbrace{\mathbb{E}_{\theta,L}\left[\bar{p}_j \cdot \Delta\bar{U}_j\right]}_{\text{better outcomes}} - \underbrace{\mathbb{E}_{L}\left[\Delta I_j\right]}_{\text{lower info cost}} . \tag{31}$$

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.

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.

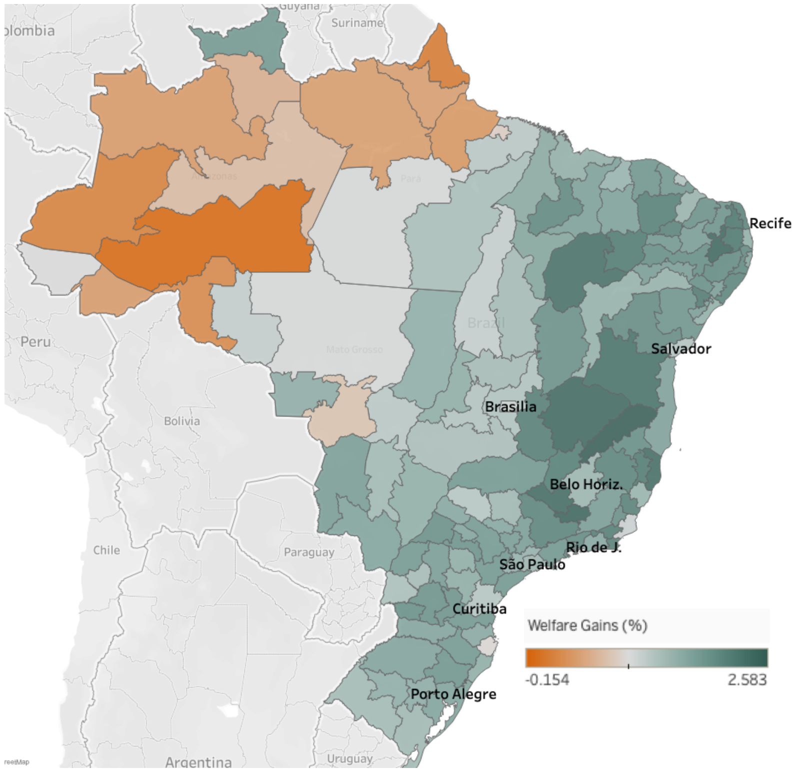
Before describing the results, it is worth emphasizing that this exercise assumes that the only effect of increased local internet penetration is to allow workers to gather information more easily. It is likely that the expansion of internet access for households has been accompanied by a parallel expansion 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 reduce information costs by a magnitude similar to the result of expanded internet access, but without affecting the local productivity process.

Figure 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 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

Figure 6: Welfare Gains by Mesoregion after Expansion of Internet Access



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 5.2—and reap the benefits in the form of high wages before more people arrive and put downward pressure on wages.

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.

References

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.

Akerman, Anders, Edwin Leuven, and Magne Mogstad, "Information Frictions, Internet and the Relationship between Distance and Trade," *Working Paper*, 2019.

Artuç, Erhan, 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., David Dorn, 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, "Hidden Income and the Perceived Returns to Migration: Experimental Evidence from Kenya," *Working Paper*, May 2019.

Bayer, Patrick, Stephen L. 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.

Beraja, Martin, Erik Hurst, and Juan Ospina, "The Aggregate Implications of Regional Business Cycles," *mimeo*, February 2019.

Bertoli, Simone, Jesús Fernández-Huertas Moraga, and Lucas Guichard, "Rational Inattention and Migration Decisions," *Working Paper*, 2019.

Blanchard, Olivier and Lawrence F Katz, "Regional Evolutions, Brooking Papers on Economic Activity," *Economic Studies Program, The Brookings Institution*, 1992, *23* (1), 1–75.

Bosch, Mariano, Edwin Goñi-Pacchioni, and William Maloney, "Trade liberalization, labor reforms and formal–informal employment dynamics," *Labour Economics*, 2012, *19* (5), 653 – 667.

Brown, Zach Y. and Jihye Jeon, "Endogenous Information Acquisition and Insurance Choice," *Working Paper*, 2019.

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.

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.

Dasgupta, Kunal and Jordi Mondria, "Inattentive importers," *Journal of International Economics*, 2018, *112*, 150 – 165.

Degroot, Morris H., "Reaching a Consensus," *Journal of the American Statistical Association*, 1974, *69* (345), 118–121.

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.

— , **Tim McQuade, and Franklin Qian**, "The Effects of Rent Control Expansion on Tenants, Landlords, and Inequality: Evidence from San Francisco," *American Economic Review*, 2019, *forthcoming.*

Dix-Carneiro, Rafael 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.

Dustmann, Christian, Albrecht Glitz, Uta Schönberg, and Herbert Brücker, "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.

Fan, Jingting, "Internal Geography, Labor Mobility, and the Distributional Impacts of Trade," *American Economic Journal: Macroeconomics*, 2019, *forthcoming.*

Farré, Lídia and Francesco Fasani, "Media exposure and internal migration — Evidence from Indonesia," *Journal of Development Economics*, 2013, *102*, 48 – 61. Migration and Development.

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 Revaled-Preference Approach to Measuring Information Frictions in Migration Decisions," *Working Paper*, June 2019.

Glitz, Albrecht and Rune Vejlin, "Learning Through Coworker Referrals," *Working Paper*, 2019.

Granovetter, Mark S., "The Strength of Weak Ties," *American Journal of Sociology*, 1973, *78* (6), 1360–1380.

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).

Kennan, John and James R. Walker, "The Effect of Expected Income on Individual Migration Decisions," *Econometrica*, 2011, *79* (1), 211–251.

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 Nordström 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.

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.

Matějka, 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.

McCauley, Jeremy, "The Role of Information in Explaining the Lack of Welfare-Induced Migration," *Working Paper*, 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.

Mian, Atif and Amir Sufi, "What Explains the 2007-2009 Drop in Employment?," *Econometrica*, 2014, *82* (6), 2197–2223.

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.

Morten, Melanie and Jaqueline Oliveira, "The Effects of Roads on Trade and Migration: Evidence from a Planned Capital City," *mimeo*, 2018.

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.

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.*

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.

Schmutte, Ian M., "Job Referral Networks and the Determination of Earnings in Local Labor Markets," *Journal of Labor Economics*, 2015, *33* (1), 1–32.

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 Matějka, "Rational Inattention Dynamics: Inertia and Delay in Decision-Making," *Econometrica*, 2017, *85* (2), 521–553.

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.

Wilson, Riley, "Moving to Jobs: The Role of Information in Migration Decisions," *Working Paper*, May 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.

Online Appendix for "Migration with Costly Information"

Charly Porcher
Princeton University

Contents

A Appendix to Section 2

A.1 Proofs

A.1.1 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 their neighbors' beliefs about any subset of states as sufficient statistics for their collective information regarding those states. The formal representation is reported in Appendix A.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.[40]

Proposition 4. *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:*

$$\log \bar{\pi}_{kt}(\theta_t) = C_{kt} + \sum_j \sum_{s_t} L_{jkt|s_t} \log \pi_{jt|s_t}(\theta_t), \tag{A.1}$$

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 $\pi = \{\pi_{j|s}\}_{s,j=1,\ldots,J} \in \Delta\Theta^{J\times S}$. Let $\bar{\Theta} = \{\theta, \hat{\theta}\}$. Denote by $\bar{\pi}(\pi) = \{\bar{\pi}_k(\pi)\}_{k=1,\ldots,J}$ 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:

$$\log \frac{\bar{\pi}_k(\pi)(\theta)}{\bar{\pi}_k(\pi)(\hat{\theta})} = g_k\left(\log \frac{\pi_0(\theta)}{\pi_0(\hat{\theta})}, \ldots, \log \frac{\pi_1(\theta)}{\pi_1(\hat{\theta})}\right). \tag{A.2}$$

[40] As shown by Levy and Razin (2018), this log-linear rule can be obtained if agents treat their marginal information sources as conditionally independent.

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 (A.2) has to be satisfied for any arbitrary pair of states, we have:

$$\begin{aligned} & g_k\left(\log \frac{\pi_0(\theta)}{\pi_0(\hat{\theta})}, \ldots, \log \frac{\pi_1(\theta)}{\pi_1(\hat{\theta})}\right) + g_k\left(\log \frac{\pi_0(\hat{\theta})}{\pi_0(\tilde{\theta})}, \ldots, \log \frac{\pi_1(\hat{\theta})}{\pi_1(\tilde{\theta})}\right) \\ = \quad & \log \frac{\bar{\pi}_k(\pi)(\hat{\theta})}{\bar{\pi}_k(\pi)(\tilde{\theta})} + \log \frac{\bar{\pi}_k(\pi)(\hat{\theta})}{\bar{\pi}_k(\pi)(\tilde{\theta})} \\ = \quad & g_k\left(\log \frac{\pi_0(\theta)}{\pi_0(\tilde{\theta})}, \ldots, \log \frac{\pi_1(\theta)}{\pi_1(\tilde{\theta})}\right) \end{aligned}$$

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 (A.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:

$$\log \frac{\bar{\pi}_k(\pi)(\theta)}{\bar{\pi}_k(\pi)(\hat{\theta})} = \sum_j \sum_s L_{jk|s} \log \frac{\pi_{jk|s}(\theta)}{\pi_{jk|s}(\hat{\theta})}. \tag{A.3}$$

A.1.2 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 (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 (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 (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 (15) given p. Together, these two relationships imply the result.

A.1.3 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 (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 (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:

$$\mathcal{B}(b_L, b_\theta, b_\varepsilon) = \left\{(L_{t-1}, \theta_t, \varepsilon_{it}) \mid L_{kt} > b_L,\ \theta_{kt} \in (-b_\theta, b_\theta),\ \varepsilon_{ikt} \in (-b_\varepsilon, b_\varepsilon),\ \forall i, k, t.\right\}.$$

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 (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.

A.1.4 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, they 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).[41]

[41] 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.

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.[42]

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 (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:*

$$\max_{q,p} \mathbb{E}\left[\sum_{t=1}^{\infty} \delta^t \left(u_{l_{t-1}l_t t}\left(\theta_t, \omega_t\right) + \lambda_{l_{t-1}}\left(\log q_{l_{t-1}l_t t}\left(\omega_t\right) - \log p_{l_{t-1}l_t t}\left(\theta_t, \omega_t\right)\right)\right)\right]. \tag{A.4}$$

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:

$$I_{l_{t-1}}\left(\omega_t\right) = \lambda_{l_{t-1}}\left(H\left(p\right) - \mathbb{E}_p\left[-\log p_{l_{t-1}l_t t}(\theta_t, \omega_t)\right]\right)$$

[42] 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}.

$$= \max_{q\in\Delta(J)} \mathbb{E}_p\left[\lambda_{l_{t-1}}\left(-\log q_{l_{t-1}l_t t}(\omega_t)+\log p_{l_{t-1}l_t t}(\theta_t,\omega_t)\right)\right]$$

Substituting this expression for the information cost into (15) gives the result in (A.4). □

I now show that (17) can be obtained as a solution to the control problem. First, I show that given any default rule q, the dynamic logit rule (17) solves problem (A.4):

Lemma 3. *Given any default rule q, the control problem with fixed q:*

$$\max_p \mathbb{E}\left[\sum_{t=1}^{\infty}\delta^t\left(u_{l_{t-1}l_t t}\left(\theta_t,\omega_t\right)+\lambda_{l_{t-1}}\left(\log q_{l_{t-1}l_t t}\left(\omega_t\right)-\log p_{l_{t-1}l_t t}\left(\theta_t,\omega_t\right)\right)\right)\right], \tag{A.5}$$

has for solution:

$$p_{jkt}(\theta_t,\omega_t)=\frac{q_{jkt}(\omega_t)\exp\left(u_{jkt}(\theta_t,\omega_t)+\delta\bar{V}_{kt+1}(\theta_t,\omega_t)\right)^{1/\lambda_j}}{\sum_l q_{jlt}(\omega_t)\exp\left(u_{jlt}(\theta_t,\omega_t)+\delta\bar{V}_{lt+1}(\theta_t,\omega_t)\right)^{1/\lambda_j}} \tag{A.6}$$

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:

$$V_{l_{t-1}t}(\theta_t,\omega_t)=\max_{p_\tau}\sum_{\tau=0}^{\infty}\delta^\tau\mathbb{E}_{\theta_\tau}\left[\sum_k p_{l_{\tau-1}k\tau}(\theta_\tau,\omega_\tau)\left(u_{l_{\tau-1}k\tau}(\theta_\tau,\omega_\tau)+\lambda_{l_{\tau-1}}\left(\log q_{l_{\tau-1}k\tau}(\omega_\tau)-\log p_{l_{\tau-1}k\tau}(\theta_\tau,\omega_\tau)\right)\right)\right]$$

The value V_{jt} satisfies the recursion:

$$V_{jt}(\theta_t,\omega_t)=\max_{p_t}\mathbb{E}_{\theta_t}\left[\sum_k p_{jkt}(\theta_t,\omega_t)\left(u_{jkt}(\theta_t,\omega_t)+\lambda_j\left(\log q_{jkt}(\omega_t)-\log p_{jkt}(\theta_t,\omega_t)\right)+\delta\bar{V}_{kt+1}(\theta_t,\omega_t)\right)\right], \tag{A.7}$$

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 (A.7), 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 (A.7) gives the recursion:

$$\bar{V}_{jt}(\theta_{t-1},\omega_{t-1}) = \mathbb{E}_{\theta_t}\left[\sum_k p_{jkt}(\theta_t,\omega_t)\left(\sum_l q_{jlt}(\omega_t)\exp\left(u_{jlt}(\theta_t,\omega_t)+\delta\bar{V}_{lt+1}(\theta_t,\omega_t)\right)^{1/\lambda_j}\right)\right],$$

$$= \quad \mathbb{E}_{\theta_t}\left[\left(\sum_l q_{jlt}(\omega_t)\exp\left(u_{jlt}(\theta_t,\omega_t)+\delta\bar{V}_{lt+1}(\theta_t,\omega_t)\right)^{1/\lambda_j}\right)\right].$$

□

To prove that the logit mobility rule in (A.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 $s=k$, $q_{jk}(L_{t-1},\varepsilon_{it})$:

$$\pi_{jk}(\theta_t) = \frac{p_{jk}(\theta_t,L_{t-1},\varepsilon_{it})\pi_j(\theta_t|L_{t-1})}{q_{jk}(L_{t-1},\varepsilon_{it})}. \tag{A.8}$$

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:

$$\log\bar{\pi}_k(\theta_t|L_t) = C_{kt} + \sum_j L_{jkt}\log\pi_{jk}(\theta_t) \tag{A.9}$$

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:

$$\pi_k(\theta_{t+1}|L_t) = \int_\theta \bar{\pi}_k(\theta|L_t)\gamma(\theta_{t+1}|\theta)d\theta, \tag{A.10}$$

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:

$$L_{kt} = \sum_j L_{jt-1}\bar{p}_{jk}(\theta_t, L_{t-1}).. \tag{A.11}$$

A.1.5 Proof of Proposition 3

After rewriting agents' problem (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:

$$\max_{\{q_{jl}\}} \int_\theta \lambda_j \log \sum_k q_{jk}(\omega_t) \exp\left(\left(\theta_k + A_k - \alpha \log L_{kt} + B_k - \kappa_{jk} + \delta \bar{V}_k(\theta, L_{t-1})\right)/\nu + \varepsilon_{ikt}\right)^{1/\lambda_j} \pi_j(\theta|L_{t-1})d\theta$$

$$s.t. \quad \sum_k q_{jk}(\omega_t) = 1, \quad q_{jk}(\omega_t) \geq 0.$$

Note that $\log \sum_k \exp(v_k/\nu) = \mathbb{E}_e[\max_k (v_k + \nu e_k)] + C$, 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:

$$\begin{aligned}
&\mathbb{E}_\theta \mathbb{E}_e \left[\max_k \left\{ \frac{A_k - \alpha \log L_{kt} + B_k - \kappa_{jk} + \delta \bar{V}_k(\theta, L_{t-1})}{\nu\lambda_j} + \log q_{jk}(\omega_t) + \frac{\varepsilon_{ikt}}{\lambda_j} + \frac{\theta_k}{\lambda_j \nu} + e_k \right\}\right] + C \\
= \; &\mathbb{E}_\theta \mathbb{E}_e \left[\max_k \left\{ \frac{\mu_{jk}(L_{t-1}) + A_k - \alpha \log L_{kt} + B_k - \kappa_{jk} + \delta \bar{V}_k(\theta, L_{t-1})}{\nu\lambda_j} + \log q_{jk}(\omega_t) + \frac{\varepsilon_{ikt}}{\lambda_j} + \underbrace{\frac{\tilde{\theta}_k}{\lambda_j \nu} + e_k}_{\equiv \zeta_j \tilde{e}_k} \right\}\right] + C \\
= \; &\mathbb{E}_{\tilde{e}} \left[\max_k \left\{ \frac{\mu_{jk}(L_{t-1}) + A_k - \alpha \log L_{kt} + B_k - \kappa_{jk} + \delta \bar{V}_k(\theta, L_{t-1})}{\nu\lambda_j} + \log q_{jk}(\omega_t) + \frac{\varepsilon_{ikt}}{\lambda_j} + \zeta_j \tilde{e}_k \right\}\right] + C \\
= \; &\log \sum_k q_{jk}(\omega_t)^{1/\zeta_j} \exp\left(\frac{\mu_{jk}(L_{t-1}) + A_k - \alpha \log L_{kt} + B_k - \kappa_{jk} + \delta \bar{V}_k(\theta, L_{t-1})}{\nu\lambda_j\zeta_j} + \frac{\varepsilon_{ikt}}{\lambda_j \zeta_j} \right),
\end{aligned}$$

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 $\text{Var}(\tilde{e}_k) = \frac{\pi^2}{6}$:

$$\zeta_j = \left(1 + \frac{6\sigma_j^2}{\pi^2 \lambda_j^2 \nu^2}\right)^{\frac{1}{2}}.$$

The last line of the previous sequence of equalities was obtained after using the property $\log \sum_k \exp(v_k/\nu) = \mathbb{E}_e[\max_k (v_k + \nu e_k)] + C$, 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:

$$\begin{aligned}
&\frac{\partial}{\partial q_{jk}(\omega_t)}\Bigg(\log \sum_k q_{jk}(\omega_t)^{1/\zeta_j} \exp\left(\frac{\mu_{jk}(L_{t-1}) + A_k - \alpha \log L_{kt} + B_k - \kappa_{jk} + \delta \bar{V}_k(\theta, L_{t-1})}{\nu \lambda_j \zeta_j} + \frac{\varepsilon_{ikt}}{\lambda_j \zeta_j}\right) \\
&+\varphi\left(1 - \sum_k q_{jk}(\omega_t)\right)\Bigg) \\
= \quad & 0,
\end{aligned}$$

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:

$$\max_k \frac{1}{\lambda_j}\left(\frac{\theta_{kt}}{\nu} + \left(A_k - \alpha \log L_{kt} + B_k - \kappa_{jk} + \delta \bar{V}_k(\theta, L_{t-1})\right)\frac{\zeta_j}{\nu(\zeta_j - 1)} + \frac{\mu_{jk}(L_{t-1})}{\nu(\zeta_j - 1)} + \varepsilon_{ikt}\frac{\zeta_j}{\zeta_j - 1}\right) + e'_{ikt},$$

where $e'_{ikt} \sim EV1$ and $\mathrm{Var}(e'_{ikt}) = \pi^2/6$. 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 $\mathrm{Var}(\tilde{e}'_{ikt}) = \pi^2/6$:

$$\varrho_j = \left(1 + \frac{\zeta_j^2}{\lambda_j^2(\zeta_j - 1)^2}\right)^{\frac{1}{2}}.$$

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 (19), define the following constants:

$$\eta_j = \frac{1}{\zeta_j} = \left(1 + \frac{6\sigma_j^2}{\pi^2 \lambda_j^2 \nu^2}\right)^{-\frac{1}{2}} \qquad \phi_j = \frac{\lambda_j \nu \varrho_j \left(\zeta_j - 1\right)}{\zeta_j} = \nu \left(1 + \lambda_j^2 \left(1 - \eta_j\right)^2\right)^{1/2}.$$

A.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.

A.2.1 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 (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_ξ^2.

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.

A.2.2 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 (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 (1). The structure of production is similar to the baseline model, with a unique freely traded

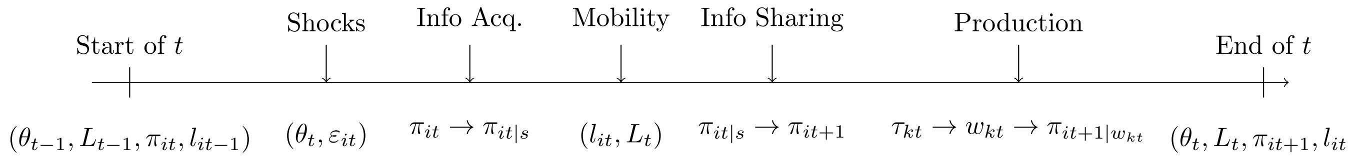


Figure A.1: Sequence of actions during any period t

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:

$$y_{kt} = \exp(A_k + \theta_{kt} + \tau_{kt}) L_{kt}^{1-\alpha}, \qquad \alpha > 0, \tag{A.12}$$

where θ_{kt} follows the AR(1) process described in (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:

$$\log w_{kt} = A_k + \theta_{kt} + \tau_{kt} - \alpha \log L_{kt} + \log(1-\alpha). \tag{A.13}$$

As a result, indirect utility, net of information costs introduced below, can be expressed as

$$u_{ijkt} = A_k + \theta_{kt} + \tau_{kt} - \alpha \log L_{kt} + B_k - \kappa_{jk} + \nu \varepsilon_{ikt}. \tag{A.14}$$

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:

$$\pi_{it}(\theta) = \Pr(\theta_t = \theta | i, t), \qquad \forall \theta \in \mathbb{R}^J. \tag{A.15}$$

Figure A.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 (A.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 $\sigma_{j,k}^2$, the resulting belief is expressed as

$$\theta_{kt} | w_{kt} = \hat{w}_{kt} \sim \mathcal{N}\left(\frac{\sigma_{j,k}^2\left(\hat{w}_{kt} - A_k + \alpha \log L_{kt} - \log(1-\alpha)\right) + \sigma_\tau^2 \mu_{j,k}}{\sigma_{j,k}^2 + \sigma_\tau^2}, \left(\sigma_{j,k}^{-2} + \sigma_\tau^{-2}\right)^{-1}\right). \tag{A.16}$$

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

$$\pi_j(\theta_{kt} | L_{t-1}, \theta_{jt-1}, \tau_{jt-1})$$

The information acquisition problem is unchanged, and agents pay the information costs (9). The updating of beliefs after having the signal $s = k$ is given by

$$\pi_{jk}(\theta_t) = \frac{\bar{p}_{jk}(\theta_t, L_{t-1}, \theta_{jt-1}, \tau_{jt-1})\pi_j(\theta_t|L_{t-1}, \theta_{jt-1}, \tau_{jt-1})}{q_{jk}(L_{t-1}, \theta_{jt-1}, \tau_{jt-1})}. \tag{A.17}$$

The population in any location k after mobility has occurred is then $L_{kt} = \sum_j L_{jkt}$, where

$$L_{jkt} = L_{jt-1}\bar{p}_{jk}(\theta_t, L_{t-1}, \theta_{jt-1}, \tau_{jt-1}), \tag{A.18}$$

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,

$$\log \bar{\pi}_k(\theta_t|L_t) = C_{kt} + \sum_j L_{jkt} \log \pi_{jk}(\theta_t). \tag{A.19}$$

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 5. *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*

$$p_{jk}(\theta_t, L_{t-1}, \theta_{jt-1}, \tau_{jt-1}, \varepsilon_{it}) = \frac{q_{jk}(L_{t-1}, \theta_{jt-1}, \tau_{jt-1}, \varepsilon_{it}) \exp\left(u_{jk}(\theta_t, L_{t-1}, \varepsilon_{it}) + \delta\bar{V}_k(\theta_t, L_{t-1})\right)^{1/\lambda_j}}{\sum_l q_{jl}(L_{t-1}, \theta_{jt-1}, \tau_{jt-1}, \varepsilon_{it}) \exp\left(u_{jl}(\theta_t, L_{t-1}, \varepsilon_{it}) + \delta\bar{V}_l(\theta_t, L_{t-1})\right)^{1/\lambda_j}} \tag{A.20}$$

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

$$V_j(\theta_t, L_{t-1}, \theta_{jt-1}, \tau_{jt-1}, \varepsilon_{it}) = \lambda_j \log\left(\sum_l q_{jl}(L_{t-1}, \theta_{jt-1}, \tau_{jt-1}, \varepsilon_{it}) \exp\left(u_{jl}(\theta_t, L_{t-1}, \varepsilon_{it}) + \delta\bar{V}_l(\theta_t, L_{t-1}, \theta_{jt-1}, \tau_{jt-1})\right)^{1/\lambda_j}\right), \tag{A.21}$$

and population and beliefs follow the laws of motion (12) - (14).

Proof. See Appendix A.1.4. □

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

$$\bar{p}_{jk}(\theta_t, L_{t-1}, \theta_{jt-1}, \tau_{jt-1}) = \frac{\exp\left(\eta_j \chi_{jk}(\theta_t, L_{t-1}, \theta_{jt-1}, \tau_{jt-1}) + \bar{v}_{jk}(\theta_t, L_{t-1}, \theta_{jt-1}, \tau_{jt-1})\right)^{\frac{1}{\phi_j}}}{\sum_l \exp\left(\eta_j \chi_{jl}(\theta_t, L_{t-1}, \theta_{jt-1}, \tau_{jt-1}) + \bar{v}_{jl}(\theta_t, L_{t-1}, \theta_{jt-1}, \tau_{jt-1})\right)^{\frac{1}{\phi_j}}}, \tag{A.22}$$

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

$$V_j(\theta_t, L_{t-1}, \theta_{jt-1}, \tau_{jt-1}) = \phi_j \log\left(\sum_l \exp\left(\eta_j \chi_{jl}(\theta_t, L_{t-1}, \theta_{jt-1}, \tau_{jt-1}) + \bar{v}_{jl}(\theta_t, L_{t-1}, \theta_{jt-1}, \tau_{jt-1})\right)^{\frac{1}{\phi_j}}\right), \tag{A.23}$$

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:

$$\bar{p}_{jk}(\theta_t, L_{t-1}) = \frac{\exp\left(\bar{v}_{jk}(\theta_t, L_{t-1})\right)^{1/\nu}}{\sum_l \exp\left(\bar{v}_{jl}(\theta_t, L_{t-1})\right)^{1/\nu}}. \tag{A.24}$$

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

$$\bar{p}_{jk}(\theta_t, L_{t-1}, \theta_{jt-1}, \tau_{jt-1}) = \frac{\exp\left(\rho_j \chi_{jk}(\theta_t, L_{t-1}, \theta_{jt-1}, \tau_{jt-1}) + \bar{v}_{jk}(\theta_t, L_{t-1}, \theta_{jt-1}, \tau_{jt-1})\right)^{\frac{1}{\psi_j}}}{\sum_l \exp\left(\rho_j \chi_{jl}(\theta_t, L_{t-1}, \theta_{jt-1}, \tau_{jt-1}) + \bar{v}_{jl}(\theta_t, L_{t-1}, \theta_{jt-1}, \tau_{jt-1})\right)^{\frac{1}{\psi_j}}}, \tag{A.25}$$

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 (A.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,

$$\frac{\text{Cov}\left(\tilde{w}_{kt}, \tilde{w}_{kt-2}\right)}{\text{Cov}\left(\tilde{w}_{kt}, \tilde{w}_{kt-1}\right)} = \rho,$$

$$\frac{1 - \rho^2}{\rho} \text{Cov}\left(\tilde{w}_{kt}, \tilde{w}_{kt-1}\right) = \sigma_\xi^2,$$

$$\text{Var}\left(\tilde{w}_{kt}\right) - \frac{\sigma_\xi^2}{1 - \rho^2} = \sigma_\tau^2.$$

A.2.3 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 2.4, 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 5.1.

A.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:

$$p_{jk}(\theta^{'i}, L^i) = g_{jk,L}(L^i)\prod_l g_{jk,l}(\theta_l^{'i})$$

$$\pi_{jk}(\theta^{'i}) = \prod_l \pi_{jk,l}(\theta_l^{'i})$$

$$V_j(\theta^{'i}, L^i) = V_{j,L}(L^i)\prod_l V_{j,l}(\theta_l^{'i})$$

Moreover, we assume that the l-partial beliefs $\pi_{jk,l}(\theta_l^{'i})$ are pdfs of normal distribution:

$$\pi_{jk,l}(\theta_l^{'i}) = \frac{1}{\sqrt{2\pi\sigma_{jk,l}^2}}\exp\left(-\frac{\left(\theta_l^i - \mu_{jk,l}\right)^2}{2\sigma_{jk,l}^2}\right)$$

and define:

$$V_{k,L}(L^{'i}) = \sum_{m,j}\beta_{k,mj}^{V,L}\log L_{mj}^i, \qquad V_{k,l}(\theta_l^{'i}) = \beta_{1k,l}^V\theta_l^{'i},$$

$$g_{jk,L}(L^i) = \exp\left(\sum_{m,j}\beta_{jk,mj}^{g,L}\log L_{mj}^i\right), \qquad g_{jk,l}(\theta_l^{'i}) = \exp\left(\beta_{1jk,l}^g\theta_l^{'i} + \beta_{2jk,l}^g\left(\theta_l^{'i}\right)^2\right).$$

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 $N = N_sN_l$ states in total, and we denote by i the index of one such state.

1. Guess partial value functions $V_{j,L}^{(0)}$, $V_{j,l}^{(0)}$, a partial choice rule $g_{jk,L}^{(0)}$, $g_{jk,l}^{(0)}$, and partial beliefs $\pi_{jk,l}^{(0)}$. Use the result of last loop if there is one. Compute the next period population that would arise from any pair $(\theta^{'i}, L^i)$:

   $$L_{jk}^{'i} = p_{jk}(\theta^{'i}, L^i)L_j^i.$$

   Define the gross payoff, exponential total payoff, and shifted exponential payoff of action $j \to k$:

   $$u_{jk}(\theta^{'i}, L^i) = A_k + \theta_k^{'i} - \alpha\log L_k^{'i} + B_k - \kappa_{jk}$$

2. Compute the end-of period homogenized beliefs for people in j that the current state is $\theta^{'i}$ if the end-of-period population is L^i:

   $$\bar{\pi}_j(\theta^{'i}|L^i) = \prod_{l=1}^J \bar{\pi}_{j,l}(\theta_l^{'i}|L^i),$$

   and the l-partial homogenized belief takes the form:

   $$\bar{\pi}_{j,l}(\theta_l^{'i}|L^i) \quad = \quad \frac{1}{C_{j,l}(L^i)}\prod_{m=1}^J \pi_{mj,l}(\theta_l^{'i})^{\frac{L_{mj}^i}{L_j^i}}$$

$$= \frac{1}{\sqrt{2\pi\bar{\sigma}_{j,l}(L^i)^2}} \exp\left(-\frac{\left(\theta_l^{'i} - \bar{\mu}_{j,l}(L^i)\right)^2}{2\bar{\sigma}_{j,l}(L^i)^2}\right)$$

with

$$\bar{\sigma}_{j,l}(L^i)^2 = \left(\frac{1}{L_j^i}\sum_{m=1}^{J}\frac{L_{mj}^i}{\sigma_{mj,l}^2}\right)^{-1}, \qquad \bar{\mu}_{j,l}(L^i) = \bar{\sigma}_{j,l}(L^i)^2 \frac{1}{L_j^i}\sum_{m=1}^{J}\frac{L_{mj}^i}{\sigma_{mj,l}^2}\mu_{mj,l}.$$

Then compute the belief for people in j that the *next* period state is $\theta^{'i}$ 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 $(\theta^{'i}, L^i)$:

$$\begin{aligned} EV_k(\theta^{'i}, L^i) &= \mathbb{E}_{\theta''}\left[V_k(\theta'', L^{'i})|\theta^{'i}\right] \\ &= V_{k,L}(L^{'i}) + \sum_l \beta_{1k,l}^V \int_{-\infty}^{\infty}\theta_l\phi_\varepsilon(\theta_l - \rho\theta_l^{'i})d\theta_l \\ &= V_{k,L}(L^{'i}) + \rho\sum_l \beta_{1k,l}^V\theta_l^{'i} \end{aligned}$$

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:

$$g_{jk,L}(L^i) = \exp\left(\sum_{m,j} \beta^{g,L}_{jk,mj} \log L^i_{mj}\right), \qquad g_{jk,l}(\theta'^i_l) = \exp\left(\beta^g_{1jk,l}\theta'^i_l + \beta^g_{2jk,l}\left(\theta'^i_l\right)^2\right).$$

6. Update the partial beliefs

$$\begin{aligned} \pi_{jk,l}(\theta'^i_l) &= g_{jk,l}(\theta'^i_l) \prod_{i=1}^{N_l} E\bar{\pi}_{j,l}(\theta'^i_l | L^i)^{\frac{1}{N_l}} \\ &= \frac{1}{\sqrt{2\pi\sigma'^2_{jk,l}}} \exp\left(-\frac{\left(\theta'^i_l - \mu'_{jk,l}\right)^2}{2\sigma'^2_{jk,l}}\right) \end{aligned}$$

with the updated mean and variances

$$\sigma'^2_{jk,l} = \left(\frac{1}{N_l}\sum_{i=1}^{N_l} \frac{1}{\hat{\sigma}_{jl}(L^i)^2} - 2\beta^g_{2jk,l}\right)^{-1} \qquad \mu'_{jk,l} = \sigma'^2_{jk,l}\left(\beta_{1jk,l} + \frac{1}{N_l}\sum_{i=1}^{N_l} \frac{\hat{\mu}_{j,l}(L^i)}{\hat{\sigma}_{jl}(L^i)^2}\right)$$

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 $\pi_{jk,l}$.

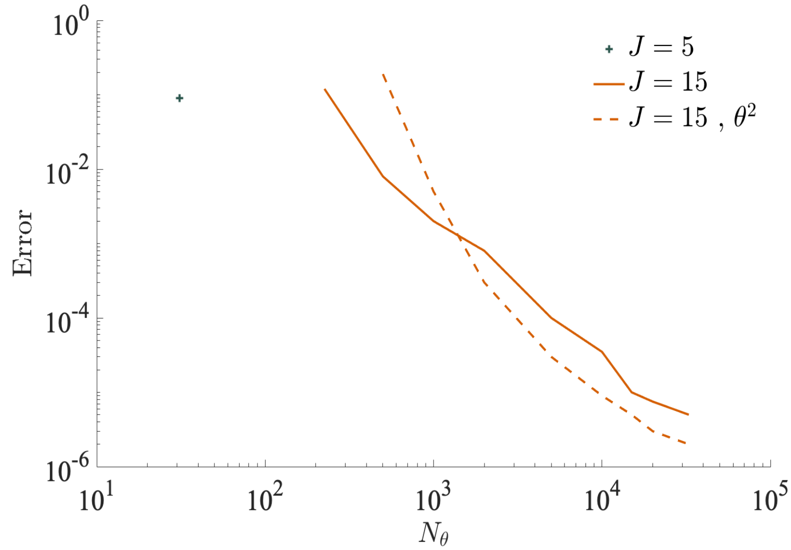
A.4 Accuracy of the Solution Algorithm

A.4.1 Binary productivity process

For few regions and a simple productivity process, it is possible to compute an "exact" solution of the model.

Figure A.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 A.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 A.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.

A.4.2 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 $L^i = (L^i_{jk})_{j,k=1,\ldots J}$ for $i = 1, \ldots, N_L$, 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 $\pi_j^{(0)}$ 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

$$\pi_j^{(0)}(\theta^{i'}|L^i) = \prod_k P^*(\theta_k^{i'}),$$

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 $p_{jk}^{(n)}$ and beliefs $\pi_j^{(n)}$, construct the ex-ante moving probabilities

$$q_{jk}^{(n)}(L^i) = \sum_{i'} p_{jk}^{(n)}(\theta^{i'}, L^i) \sum_{i''} \pi_j^{(n)}(\theta^{i''}|L^i)\Gamma(\theta^{i'}|\theta^{i''})$$

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 Appendix to Section 3

B.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.[43]

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.[44]

B.2 Additional Descriptive Statistics on Migration Patterns

In this section, I describe additional statistics about the sample. As illustrated by Figure B.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 illustrates a steep decrease of the migration rate with age. Figure B.3 shows that females in the sample are about half as mobile as males. Figure B.5 illustrates that the migration rate is increasing for highly educated, decreasing for low-educated.

Table B.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
	Total 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
	Total Income	22,976	64,657	0.00	6,068,441	746	5,443	51,529
Average	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
	Total Income	18,822	54,226	0.00	6,068,441	425	3,681	45,696

[43] 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.

[44] 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, 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.1: Migration rates across mesoregions by Year over 2000-2014

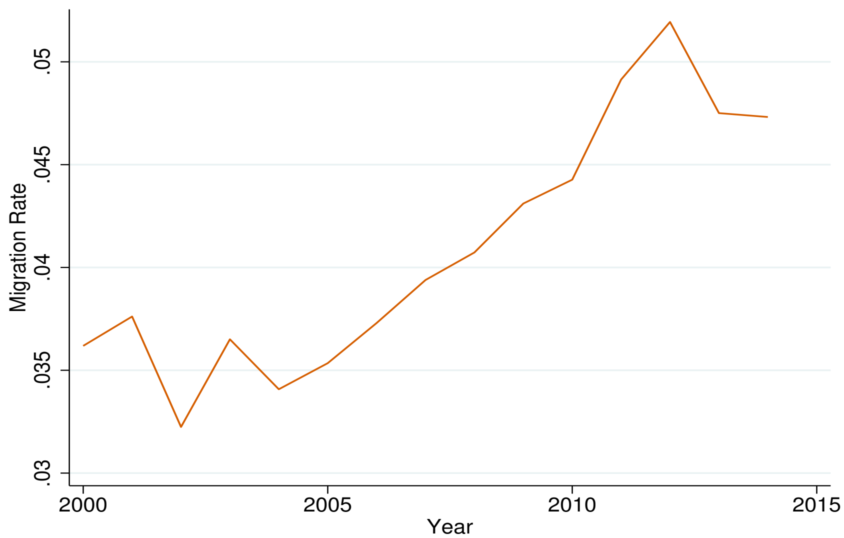


Figure B.2: Migration rates across mesoregions by Year, by age, over 2000-2014

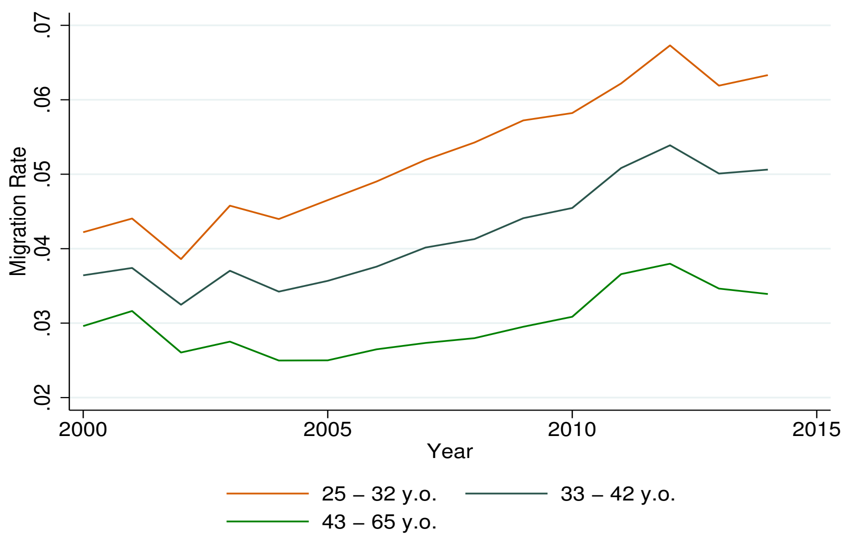


Table B.1 provides descriptive information on the set of workers who appear at least once in the RAIS data between January 2000 and December 2014. The average worker in our sample (weighted by months in sample) is 39.4 years old and has annualized earnings of R$ 18,822 (in 2014 R$), before taxes. Over 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 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.3: Migration rates across mesoregions by Year, by gender, over 2000-2014

Figure B.4: Migration rates across mesoregions by Year, by race, over 2000-2014

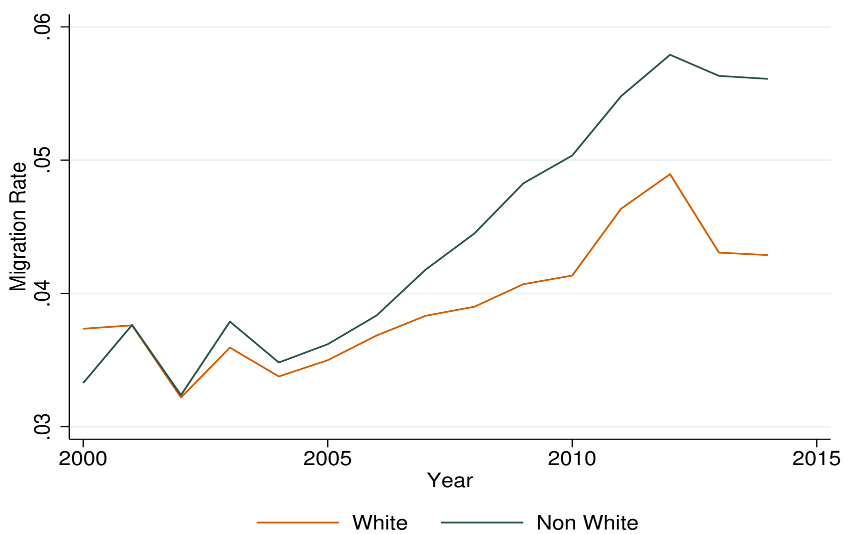
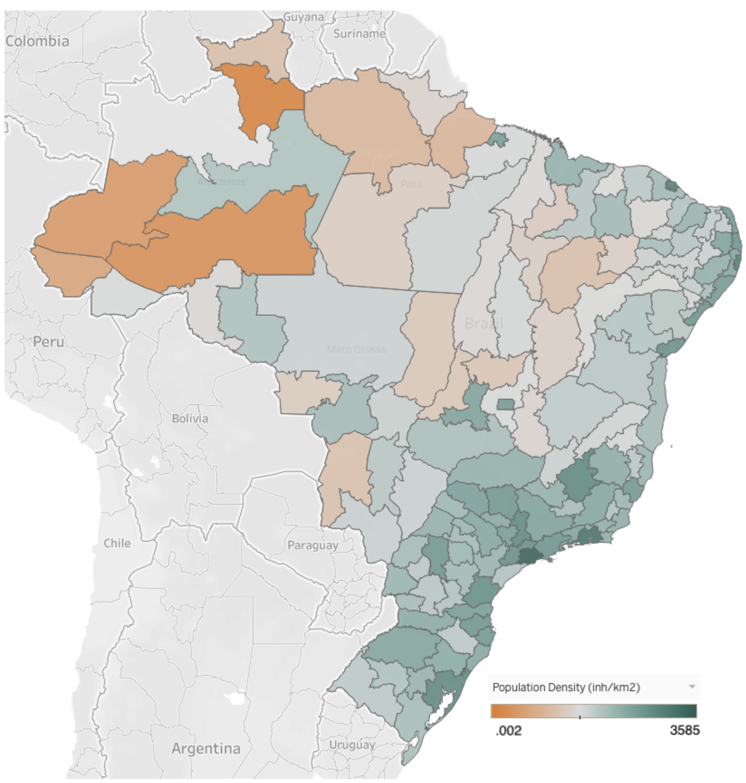


Figure B.5: Migration rates across mesoregions by Year, by educational attainment, over 2000-2014

Figure B.6: Population Density by Mesoregion - Average over 2000-2014



C Appendix to Section 4

C.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 $\{A_k\}_{k=1,\ldots J}$
  - Regional constant amenities $\{B_k\}_{k=1,\ldots J}$
- The discount rate is set to $\delta = 0.96$.
- From the wage equation, we can estimate $\{A_k\}_{k=1,\ldots J}$, 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

$$\Lambda_{jklt} = \log \left(\frac{p_{jkt} p^{\delta}_{klt+1}}{p_{jjt} p^{\delta}_{jlt+1}} \right), \qquad Q_{jklt} = \log \left(\frac{q_{jkt} q^{\delta}_{klt+1}}{q_{jjt} q^{\delta}_{jlt+1}} \right)$$

$$\Delta w_{jkt} = w_{kt} - w_{jt},$$

$$K_{jkl} = \kappa_{jk} + \delta (\kappa_{kl} - \kappa_{jl})$$

$$\chi_{jkt} = EV_{kt,t+1} - V_{kt+1} - (EV_{jt,t+1} - V_{jt+1})$$

$$EV_{kt,t+1} = \mathbb{E} [V_k(\theta_{t+1}, L_t) | \theta_t, L_{t-1}]$$

- In the data, I observe Λ_{jklt}, Δw_{jkt} and $dist_{jk}$ for all regions and years, as well as the population distribution $\{L_{jkt}\}_{j,k=1,\ldots J}$. 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 $(\lambda^0_l, \lambda^0_h)$, (ϕ^0_0, ϕ^0_1) and $\{B^0_k\}_{k=1,\ldots J}$.
  2. Simulate the model using these parameters and the previously estimated parameters.
  3. For any triplet-year $jklt$, compute the simulated Q^0_{jklt}, K^0_{jkl} and χ^0_{jkt}, and run the regression:

$$\Lambda_{jklt} - Q^0_{jklt} = \beta_1 (\Delta w_{jkt} + \delta \chi_{jkt}) + \beta_2 K^0_{jkl} + D_k - D_j$$

  4. Update $\lambda^1 = \beta_1^{-1}$, $\phi^1 = \beta_2 / \beta_1$, $B^1_k = D_k$, and return to step 1 until convergence.
- Joint estimation:

– For any $(\lambda_l^0, \lambda_h^0)$, (ϕ_0^0, ϕ_1^0) and $\{B_k^0\}_{k=1,\ldots J}$, define

$$\varepsilon_{jklt} = \Lambda_{jklt} - \left(Q_{jklt} + \frac{1}{\lambda_j} \Delta w_{jkt} - \frac{1}{\lambda_j} K_{jkl} + \frac{1}{\lambda_j} (B_k - B_j) + \frac{\delta}{\lambda_j} \chi_{jkt} \right).$$

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:

$$\mathbb{E}\left[\varepsilon_{jklt}\right] = 0,$$

$$\mathbb{E}\left[\varepsilon_{jklt} Q_{jklt}\right] = 0,$$
$$\mathbb{E}\left[\varepsilon_{jklt} \Delta w_{jkt}\right] = 0,$$
$$\mathbb{E}\left[\varepsilon_{jklt} K_{jkl}\right] = 0,$$
$$\mathbb{E}\left[\varepsilon_{jklt} (B_k - B_j)\right] = 0,$$
$$\mathbb{E}\left[\varepsilon_{jklt} \chi_{jkt}\right] = 0,$$

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

C.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 C.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.

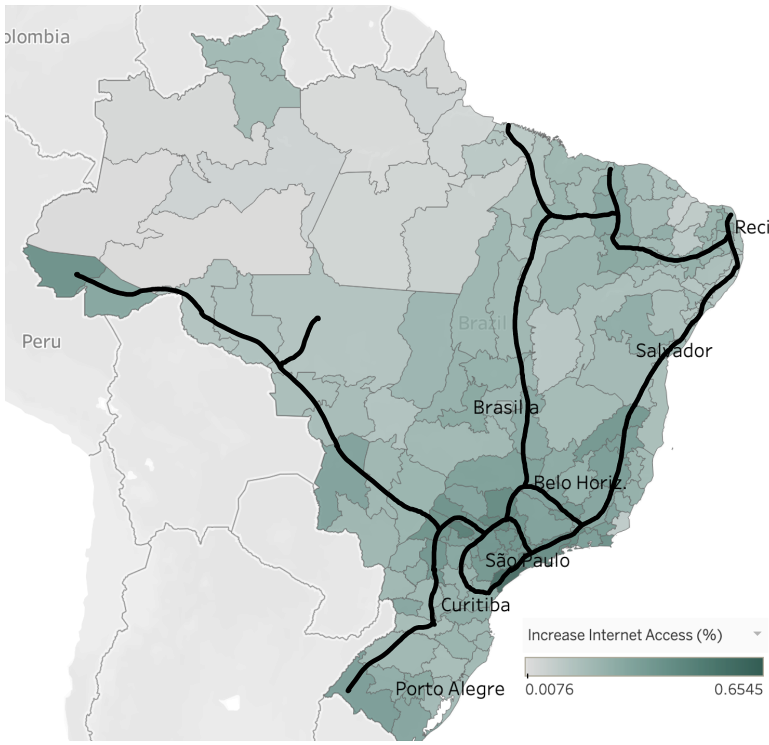
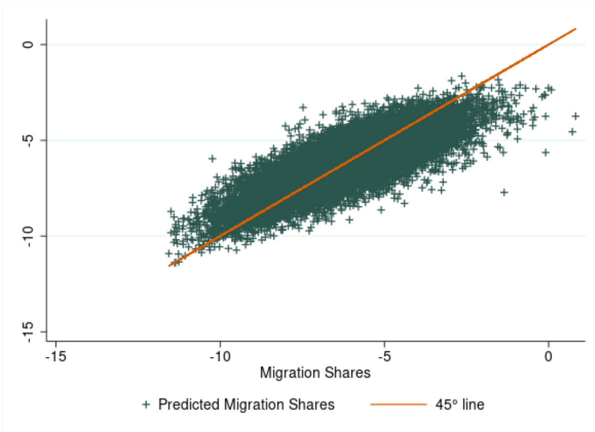


Figure C.1: Change in internet access from 2000-2007 to 2008-2014 and new backbone cables

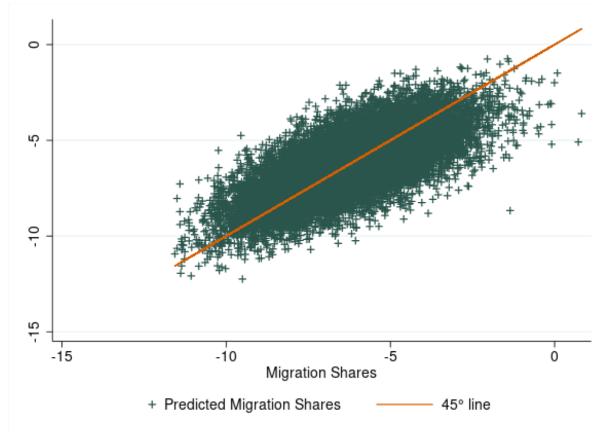
D Appendix to Section 5

D.1 Fit of Migration Flows

In Figure D.1a and Figure D.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

D.2 Heterogeneous Migration Elasticities in the Model with No Information Frictions

D.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.

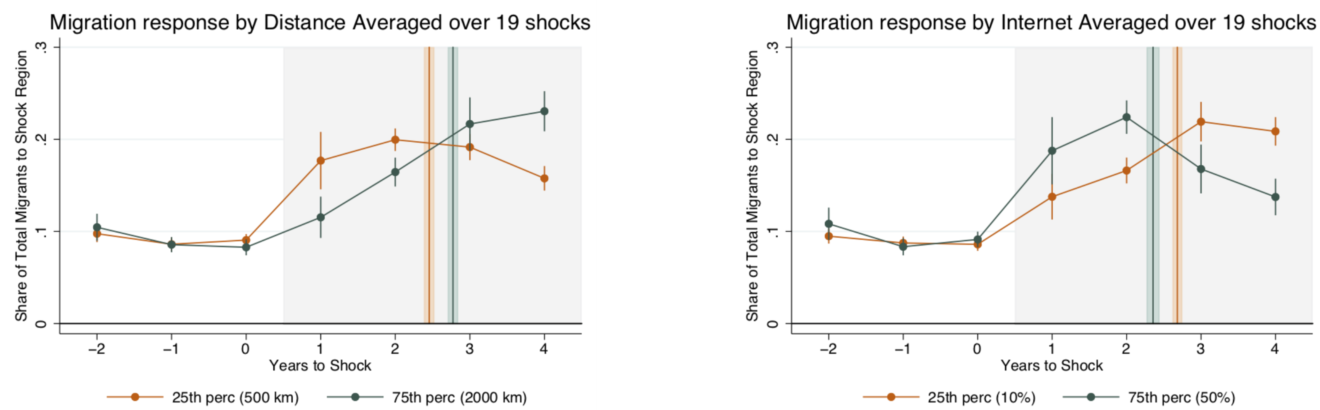
Given that the last year for which I observe migration decisions in the data is 2014 and that I want to study the evolution of the migration flows to a region experiencing a positive labor demand shock in the four years subsequent to the shock, I only search for regions that experienced positive labor 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 D.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 D.1: Description of Positive Local Labor Demand Shocks

Municipality	State	Year Shock	Number Migrants	Source Labor 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

D.4 Replicating Local Shocks in the Model

Figure D.2: Average Delay in the Migration Response to the 10 Local Shocks



(a) Distance: Data

(b) Internet: Data

Figure D.2a and Figure D.2b illustrate the estimates that arise from pooling the data across the ten shock listed in Table D.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 $\{\gamma^k_{it+s}; k = 1, \dots, K\}$ 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 D.2a and D.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.

E Additional Figures

E.1 Cardell Distribution

$$g_\beta(z) = \frac{1}{\beta}\sum_{n=0}^{\infty}\frac{(-1)^n\, e^{-nz}}{n!\Gamma(-\beta n)}.$$

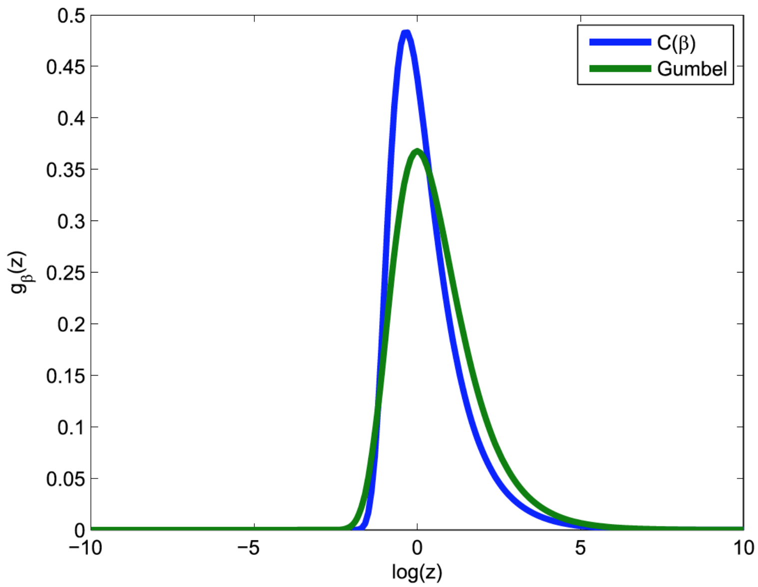


Figure E.1: Distribution Cardell with dispersion $\beta = 0.5$, vs. Gumbel. Source: Dasgupta and Mondria (2018).