Migration with Costly Information

Charly Porcher*

Princeton University

cporcher@princeton.edu

Job Market Paper

Download Most Recent Version Here

October 20, 2019

Abstract

Information is important for migration decisions. How costly is information acquisition, how do these costs depend on the economic environment, how do they affect migration decisions, and ultimately, the spatial allocation of workers? To investigate these questions, I develop a quantitative dynamic model of migration with costly information acquisition and local information sharing. Agents are rationally inattentive and migration shares feature a tractable logit structure. Information frictions affect both the magnitude and the responsiveness of migration flows to variations in local opportunities. I apply the model to internal migration in Brazil and estimate it using migration flows between 137 regions. Finding where to go appears to be as costly as moving, with annualized average costs of information and migration both equal to 3% of earnings. The model successfully predicts the observed heterogeneous migration elasticities and delayed migration responses to local shocks. To illustrate its quantitative implications, I evaluate the counterfactual effect of the roll-out of broadband internet in Brazil. By allowing workers to make better mobility choices, expanding internet access increases average welfare by 1.6%, reduces migration flows by 1.2% and reduces the cross sectional dispersion in earnings by 4%.

*Department of Economics, Princeton University, Julis Romo Rabinowitz Building, Princeton, NJ 08544. 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, 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

When local economic opportunities arise, labor mobility can help distribute gains across regions. [1] Yet, in many occasions, workers' migration patterns fail to respond to these regional shocks. A common explanation for this lack of response is the existence of large frictions to the mobility of workers. Such large migration frictions in turn have been shown to have important implications for aggregate labor productivity and welfare.[2] One explanation for workers' limited migration responses may be their lack of information about the potential gains from migration. In this paper, I assess the scope for improving the spatial allocation of workers by expanding their access to information.

Tackling this question requires to determine how much workers know about the opportunities available in other regions. Workers likely obtain information both by collecting it individually and through social interactions within local networks. Since mobility patterns shape local networks, the structure of information itself might be affected by migration. There may also be geographic disparities in access to information, with richer regions offering better infrastructure in the form of internet coverage for example. Therefore, migration opportunities themselves and what workers know about them are likely to depend both on where workers reside and who they interact with. This calls for considering the structure of information as an outcome determined in equilibrium along with migration patterns and the spatial distribution of earnings.

My main contribution is to propose a quantitative dynamic model of migration with endogenous incomplete information. In the model, the structure of information is an equilibrium outcome jointly determined with migration and earnings. Every period, unobserved region-specific productivity shocks alter the spatial distribution of payoffs. Agents are rationally inattentive and can acquire information at some cost to refine their beliefs of the productivity shocks in every region. This cost is allowed to vary by region. Agents then use their beliefs to make location decisions, facing fixed bilateral costs of moving between any two regions. Every period, the information acquired individually is shared between agents in the same region. In a given region, agents are heterogeneous in two dimensions: first, they may have different beliefs about the distribution of payoffs in each region and second, they may have different preferences for each region. The combination of these two sources of heterogeneity leads to migration patterns that feature gross flows between regions. While preference-based migration reflects utility maximization, migration decisions under incomplete information are prone to mistakes.

The model allows for a tractable and intuitive solution in stochastic steady state that delivers three new 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

[1] Labor demand shocks differ widely across regions within countries (Bartik, 1991; Blanchard and Katz, 1992). This cross-regional heterogeneity in labor demand shocks may be due to regions' differential exposure to international trade competition (Topalova, 2010; Kovak, 2013; Autor et al., 2013; Adão, 2016; Dix-Carneiro and Kovak, 2017; Pierce and Schott, 2018; Dix-Carneiro and Kovak, 2019) and changes in technology (Beaudry et al., 2010; Autor and Dorn, 2013; Bustos et al., 2016; Acemoglu and Restrepo, 2017; Hombert and Matray, 2019). Cross-regional differences in housing net worth (Mian and Sufi, 2014), government spending (Nakamura and Steinsson, 2014), business cycles (Beraja et al., 2019), the location of firms (Greenstone et al., 2010), and natural resource discovery (Feyrer et al., 2017) have also been documented as sources of regional dispersion in labor demand shocks.

[2] For evidence on the lack of labor mobility in reaction to regional labor demand shocks, see Adão et al. (2019); Yagan (2019) and several of the papers cited in footnote 1. For quantifications of the aggregate implications of workers' lack of labor mobility, see Morten and Oliveira (2018); Bryan and Morten (2018); Caliendo et al. (2018, 2019).

reduces to existing logit models of migration driven by preference heterogeneity.[3] Second, agents optimally choose to be more informed about regions offering high average payoffs, leading their migration decisions to feature predispositions towards nearby regions. By enhancing flows to closer destinations at the expense of more remote ones, these predispositions act as endogenous bilateral migration costs. Third, the model disentangles the contribution of idiosyncratic preferences and of the lack of information to migration decisions. Although both tend to reduce the responsiveness of migration flows to variations in earnings, information frictions particularly affect the response to unobserved payoffs. Recovering the elasticity of migration with respect to unobserved productivity shocks in addition to the elasticity with respect to wages allows to separately identify information frictions from preference heterogeneity.

In the second part of the paper, I apply the model to internal migration in Brazil in order to assess the quantitative importance of information frictions in this context. Although the model is analytically tractable and delivers closed form expressions, simulating it on a realistic scale requires to confront a severe curse of dimensionality. Indeed, the state space comprises the vector of productivity and population in each region, as well as the beliefs inherited from previous periods. To overcome this challenge, I first show that in stochastic steady state, the beliefs can be described as a function of the productivity and population only, reducing drastically the effective state space. 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 and approximate equilibrium beliefs by a specific distribution.[4] I confirm the accuracy of the algorithm by comparing it to an almost-exact solution when the number of regions is small or when the productivity process is discrete.

The tractable structure of the model makes it amenable to quantification. To estimate the relevant parameters of the model, in particular the dispersion of preferences and the cost of information acquisition, I rely on detailed migration flows between the 137 Brazilian regions that map the whole country over 15 years. I construct these flows from administrative matched employer employee data covering the universe of workers employed in the formal sector for the period 2000 to 2014. I observe workers' location and earnings every year, representing more than 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. Migration flows vary differently in response to changes in payoffs depending upon whether agents directly observe these payoffs or if they must acquire information about them. A higher cost of information acquisition leads to less precise beliefs about unobserved payoffs and a lower responsiveness of migration flows.

The estimated cost of information amounts on average to 3% of earnings being paid every year to acquire information about other regions. The estimated information costs are lower in regions with a higher fraction of residents with an internet connection, as well as in regions with higher wages and higher population density. The annualized average bilateral migration cost is also 3% of earnings, so that finding where to go appears as costly as moving. Importantly, the estimated migration costs are 40% smaller than what we would obtain if we assumed that information costs were zero. This is due to the omitted effect of endogenous predispositions.

I then show the estimated model successfully predicts two key features of observed migration patterns

[3]See for example Caliendo et al. (2019).

[4]I assume that beliefs can be approximated by a Cardell distribution, see Cardell (1997); Brown and Jeon (2019).

that the model with complete information cannot. First, 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. In the model, this higher responsiveness of mobility decisions arises because agents' beliefs are more tightly correlated with the true payoffs in regions with which they have more interactions, and if they can acquire information at a lower cost. Second, in response to a positive local shock in a region, the migration response to this region is slower for origins that are more distant or have lower internet access. In the model, this delay is due to the gradual updating of beliefs about the shock thanks to local information sharing. In regions more distant or with higher information costs, this local information sharing is slower.

Finally, I illustrate the quantitative implications of the model by undertaking two counterfactual exercises. First, I evaluate the effect of removing all information frictions by comparing the steady state equilibrium with the estimated information costs to the one in which these costs are set to zero. I find that welfare increases by 5.5%, with a 7% decrease in the cross-sectional dispersion in earnings reflecting a better arbitrage of local shocks. In the steady state with complete information, gross migration flows are more concentrated towards high payoffs regions and are lower by 4.1%.

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 one in which the change in information costs reflects the observed internet access in 2014. I find that the expansion of internet access increases average welfare by 1.6%. I then decompose these welfare changes into several sources. Some gain is mechanically coming from the decrease in the information cost. A second channel is the change in the information transmitted by local networks so that workers may have to spend less on individual information acquisition. A third potential source of gains is the better sorting of agents to regions offering higher payoffs. Finally, some gains may arise from a change in regional outcomes due to a response of wages and future values of residing in a location. The positive average welfare gains mask substantial heterogeneity in welfare effects, with several remote regions experiencing small losses despite obtaining some internet access. This is because fewer well-informed workers move to these relatively unattractive regions in a steady state with better information access, making local information sharing less effective, and ultimately leading to a worsening in workers' spatial sorting.

The rest of the paper is organized as follows. Section 3 describes the data and the empirical evidence. Section 2 describes the model, Section 4 describes the estimation, Section 5 confronts some of the model's predictions to the data, Section 6 presents the counterfactual exercises while Section 7 concludes.

This paper is related to several existing literatures. A recent set of empirical studies emphasizes the importance of information frictions in 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 sources, arising either from differential media exposure (McCauley, 2019; Farré and Fasani, 2013; Wilson, 2018), or randomized treatment (Baseler, 2019; Bryan et al., 2014).[6]

[5] I abstract from the effects that broadband internet access may have had on local productivity and focus only on its role in improving migrants' information.

[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

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 showed that they were not aligned 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) show that migration patterns are not consistent with complete information. They find that migrants' information is concentrated on 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 which leads to an endogenous structure of information that is also concentrated on nearby regions and larger cities. This framework then allows me to quantify the implications of these information frictions for aggregate welfare.

My paper also contributes to the economic geography literature focused on identifying the determinants of labor mobility. In their analysis of internal migration patterns in the United States, Kennan and Walker (2011) pointed out that large migration costs were 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.[7] 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.

By incorporating the role of information frictions in a dynamic spatial equilibrium, this paper also contributes to the quantitative economic geography literature. Recent analyses have emphasized the relevance of spatial linkages due to trade and labor mobility for the adjustment of economies to various shocks (Monte et al., 2018; Tombe and Zhu, 2019; Caliendo et al., 2019). I describe how the endogenous structure of information interacts with the spatial allocation of economic activity, creating an additional channel of adjustment with important welfare implications.

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

Finally, this paper is related to the emerging 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 multinomial logit rule.[8] This property was then extended to single-agent dynamic problems by Steiner et al. (2017), 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-

outcomes at a finer geographical level.

[7]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.

[8]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.

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 of Eaton and Kortum (2002).[9] I contribute to this literature by providing the first structural estimation of a dynamic rational inattention model, and show in particular how to disentangle preferences from beliefs using observed choices and payoffs.

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. I first describe the environment and the structure of flow payoffs. Second, I present the individual information acquisition problem faced by rationally inattentive agents. Third, I describe the local information sharing. Fourth, 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 vary over time, and that it may be too difficult for agents to track these fluctuations perfectly. Agents are therefore faced with a trade-off between the value of holding precise information about the current payoffs, and the costs of gathering such information.

With this goal in mind, I consider an infinite-horizon environment with J regions. Each region j is characterized by a time-invariant productivity level A_j and amenities B_j. While the cross-sectional variation in baseline productivities A_j will help explain persistent spatial dispersion in earnings, the variation in amenities B_j will justify why some regions would attract more workers than others with similar earnings levels. Every period, regions experience an exogenous stochastic productivity shock θ_{jt}, meant to capture the fluctuations in migration opportunities over time.[10] Denote by $\Gamma(\theta_t|\theta^{t-1})$ 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 must be paid by agents if they decide to move from any region j to another region k.

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 every regions, capturing the fact that workers that are observationally equivalent from the point of view of the researcher would in fact benefit differently from a given migration decision. Denoting by $u(c_{kt})$ the flow of utility from consumption derived by agents

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

[10] As discussed in footnote 1, such fluctuations in local labor demand may be due to regions' 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 their analysis of migration patterns in Brazil, Fujiwara, Morales and Porcher (2019) focused on a number of such local shocks, ranging from dam construction, mining and oil booms, to surges in tourism activity.

in region k in period t, we 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, and where ε_{ikt} represents agent i's idiosyncratic taste for region k at t, scaled by the parameter ν.

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. In Appendix B.2, I extend the model to a multi-sector economy with costly interregional trade between regions where production uses labor, housing and intermediate goods as inputs, in the spirit of Caliendo et al. (2019). In this simpler version, there is a unique freely traded homogeneous good chosen as numeraire. In addition, I assume that the production function takes a Cobb-Douglas form using labor as the single input, so that the output in region k at t writes:

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

where L_{kt} is the human capital in region k at t, equal to the population in k since I assume that every agent supplies one unit of human capital. The parameter α captures the decreasing marginal product of labor, and is a source of congestion in the model that will reduce the attractiveness of a region as more workers move in. More generally, such congestion could also arise because of local congestion in housing. A positive value of α brings about general equilibrium effects in the cross-regional distribution of earnings as wages depend on the endogenous allocation of workers. Setting the constant C to equal $1/(1-\alpha)$, this becomes apparent since the wage arising from profit maximization writes in log:

$$\log w_{kt} = A_k + \theta_{kt} - \alpha \log L_{kt}. \tag{3}$$

Moreover, the indirect utility can then be expressed as:[11]

$$u_{ijkt} = \log w_{kt} + B_k - \kappa_{jk} + \nu\varepsilon_{ikt}, \tag{4}$$

Now turning to the incompleteness of information, I assume that the stochastic productivity vector θ_t is the only imperfectly observed variable. One justification for this is that time-invariant variables such as amenities, baseline productivities and amenities have been learned gradually over time until they became perfect knowledge. 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 region. In Appendix B.3, I extend the model to allow for unobserved population, so that agents hold beliefs about both productivity and population. It may also seem plausible that local amenities would also be subject to stochastic variation. In Appendix B.4, I extend the model to allow for unobserved amenity shocks so that agents would track the sum of productivity and amenity shocks. As a result, agents hold beliefs about the cross-sectional distribution of productivity. Denote by $\pi_t = (\pi_{1t}, \ldots, \pi_{Jt})$ the set of region-specific beliefs

[11] Since the production exhibits decreasing returns, firms make positive profits. Here I assume that profits are distributed to capitalists who spend the profits in consumption.

about the current distribution of productivity, so that

$$\pi_{jt}(\theta) = \Pr(\theta_t = \theta | j), \qquad \forall \theta \in \mathbb{R}^J. \tag{5}$$

Before describing how agents' beliefs are formed, I specify the process of productivity. 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 every period t, the productivity shock in region j is related to the previous period shock according to the following expression:

$$\theta_{jt} = \rho\theta_{jt-1} + \xi_{jt}, \qquad \xi_{jt} \sim \mathcal{N}\left(0, \sigma_\xi^2\right). \tag{6}$$

Assuming that θ_t follows an AR(1) recognizes that local shocks are likely to feature some persistence over several periods, so that any information acquired about past values of shocks has value for predicting current shocks. However, the AR(1) process imposes that the dependence of the current shock on past realization can be summarized by the shock in the previous period. This implies that agents only use their beliefs about the previous period productivity to form beliefs about the current distribution of shocks.

In this economy, the state variables that are common to all agents are the vectors of productivity, population inherited from the previous period, and beliefs $(\theta_t, L_{t-1}, \pi_t)$, while the states specific to each agent are their location at the beginning of the period $l_{t-1} \in \{1, \ldots, J\}$ and their idiosyncratic preference shocks ε_t. Within a period, three steps are realized sequentially. First, once the productivity vector θ_t and idiosyncratic preferences ε_t are realized, agents can refine their prior beliefs by acquiring and processing information. This step is governed by the rational inattention problem described in Section 2.2, and leads agents to form posterior beliefs about θ_t. Second, agents use their posterior beliefs to compute the expected payoffs in every region, and move to the region offering the highest expected payoff. This leads to a new distribution of population described in Section 2.3. Third, once agents reach their destination, they engage in local information sharing. After communicating their beliefs to all agents in the same region, agents reach a consensus about the distribution of θ_t. This step is described in Section 2.4. I then define the equilibrium in Section 2.5 and discuss its steady-state solution in Section 2.6.

2.2 Individual Information Acquisition

I now describe the rational inattention problem at the center of the trade off faced by agents between making accurate predictions about the payoffs in each region and paying the cost of acquiring and processing a large amount of data to make such predictions.

Following Matějka and McKay (2015) and Steiner et al. (2017), individuals in region j start with a prior π_{jt} about θ_t and rationally choose how much information to acquire about θ_t to form posterior beliefs about the payoffs in each location. To do so, they can choose to receive signals 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_{jt}, they choose the conditional distribution $f(s|\theta_t, \pi_{jt})$. Their information acquisition strategies $f(\cdot)$ are unconstrained, reflecting the idea that agents can gather information in many different ways. In particular, individuals may wish to become partially informed about regions, i.e. receive a vector of signals with limited information content. Once an agent chooses her signal structure for the period, nature draws a signal realization s from $f(\cdot)$. Given the signal, the agent updates her prior, resulting in the posterior belief $\pi_{jt|s}$ after applying Bayes' rule:

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

Although agents are free to design any signal structure, there is a cost to information acquisition so that more "informative" signal structures are more costly. I focus on entropy-based information costs.[12] 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:

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

It is a measure of uncertainty about X.[13] Starting from a prior belief distribution π_{jt}, a signal distribution which 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.[14] I define:

$$I_j\left(f|\pi_{jt}\right) = \lambda_j\left(H\left(\pi_{jt}\right) - \mathbb{E}_s\left[H\left(\pi_{jt|s}\right)\right]\right), \tag{9}$$

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 $H\left(\pi_{jt}\right)$. After, it becomes $H\left(\pi_{jt|s}\right)$. The region-specific parameter λ_j scales this information cost, reflecting that some regions may offer more efficient technologies to gather information and reduce uncertainty.

An information strategy therefore consists in assigning a signal distribution $f_t(s_t|l_{t-1}, \theta_t, L_{t-1}, \pi_{l_{t-1}t}, \varepsilon_t)$ for each origin location l_{t-1}, productivity θ_t, population distribution L_{t-1}, prior beliefs $\pi_{l_{t-1}t}$ and preference draws ε_t. Agents also devise action strategies $\sigma_t(l_{t-1}, s_t, \pi_{l_{t-1}t}, L_{t-1}, \varepsilon_t)$ indicating the mobility choice at time t for each origin location l_{t-1}, prior beliefs $\pi_{l_{t-1}t}$, current costly signal s_t, population distribution L_{t-1} and preference draws ε_t, such that agents solve the following problem:

$$\max_{f,\sigma} \mathbb{E}_{\theta^t, \varepsilon^t}\left[\sum_{t=1}^{\infty} \delta^t\left(u_{l_{t-1}l_t}(\theta_t, L_t, \varepsilon_t) - I_{l_{t-1}}(f_t|\pi_{l_{t-1}t})\right)\right] \tag{10}$$

where $l_t = \sigma_t(l_{t-1}, s_t, \pi_{l_{t-1}t}, L_t, \varepsilon_t)$ is the location optimally chosen at t, the information cost is defined in (9), and the flow payoff is given by (4), substituting wages by their expression in (3).

[12]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.

[13]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.

[14]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.

2.3 Mobility

Within period t, once agents have acquired information and formed posterior beliefs $\pi_{jt|s_t}$, they use these beliefs to compute the expected payoffs in each region. They then move to the region offering the highest expected payoffs. These mobility decisions lead to a new allocation of workers across regions, namely a new population distribution L_t. Denote by $\bar{p}_{jkt}(\theta_t, L_{t-1}, \pi_{jt}, s_t)$ the probability that an agent would move from region j to k at t after receiving the signal s_t if the current productivity is θ_t, prior beliefs in j are π_{jt}, and the population distribution is L_{t-1}. Using the optimal information f_t and mobility strategy σ_t, this probability writes:

$$\bar{p}_{jkt}(\theta_t, L_{t-1}, \pi_{jt}, s_t) = \mathbb{E}_{\varepsilon_t}\left[f_t(s_t|j, \theta_t, L_{t-1}, \pi_{jt}, \varepsilon_t)\mathbb{1}\{\sigma_t(j, s_t, \pi_{jt}, L_{t-1}, \varepsilon_t) = k\}\right], \tag{11}$$

We assume that the total population in the economy stays constant to $\bar{L}$ over time, 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 mobility has occurred is then $L_{kt} = \sum_j L_{jkt}$, where:

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

2.4 Local Information Sharing

In practice, on top of collecting information individually, 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 origin location may learn about the payoffs in other regions by interacting with newly arrived workers from these regions who are likely to have relatively accurate information about their origin. This may then influence their future location decisions. As I will show in Section **??**, migration patterns in Brazil seem to be consistent with workers relying on the information provided by their local networks for their migration decisions. To capture this fact, after they reach their new destination, agents are able to collect additional information about migration opportunities by engaging in communicational social learning with other agents in their location. Following Molavi et al. (2018), I now describe the assumptions on the behavior of agents leading to a tractable and intuitive social learning rule that aggregates the information of all agents in a region into a unique belief distribution.

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. 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. In Appendix B.5, I derive an extension in which agents place different weights on other agents depending on whether they belong to the same demographic or occupational group as them.

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 is commonly viewed as demanding unrealistic cognitive abilities on individuals.[15] Here, I postulate that agents follow social learning rules that satisfy

[15] Starting with Degroot (1974), a rich literature has proposed relatively simple functional forms on agents' learning rules,

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 1 (Imperfect Recall). *$\bar{\pi}_{kt}$ is independent of $\pi_{\tau-1}$ for all k and all $\tau \leq t-1$.*

In order to obtain a simple unique characterization of the social learning rule, I follow Molavi et al. (2018) and impose three natural additional restrictions on how agents process their neighbors' information. The first one is that agents' social learning rules are *label neutral*, which means that relabeling the underlying states has no bearing on how agents process information. Second, I assume that individuals do not discard their neighbors' most recent observations by requiring their social learning rules to be increasing in their neighbors' last period beliefs, a property referred to as *monotonicity*. Third, I require agents' learning rules to satisfy *independence of irrelevant alternatives*: each agent treats her neighbors' beliefs about any subset of states as sufficient statistics for their collective information regarding those states. The formal representation is reported in Appendix B.1.1.

Molavi et al. (2018) show that, in conjunction with imperfect recall, these three restrictions lead to a unique representation of agents' social learning rules up to a set of constants: at any given time period, the log-likelihood ratios of all agents' beliefs are combined linearly, weighted by their centrality in the network. Given the assumption that agents are all connected to each other locally, they have the same centrality and the beliefs of every agents are given the same weight.[16]

Proposition 1. *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:*

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

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 $\int_\theta \bar{\pi}_{kt}(\theta)d\theta = 1$.

Proof. See Appendix B.1.1. □

Besides its simplicity, this social learning rule has a very intuitive interpretation. First, since all agents have the same weight, a particular posterior belief $\pi_{jt|s_t}$ 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

with the objective of capturing the richness of the network interactions while maintaining analytical and computational tractability.

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

the final beliefs.[17] 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.[18]

The sharing of 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 the region of origin of agents. 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 according to the assumption that θ_t follows an AR(1) process.

2.5 Dynamic Rational Inattention Equilibrium

We are ready to define a competitive equilibrium in our economy. It 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.

Definition 1. *Given an initial population distribution L_0, initial beliefs π_0, an equilibrium is a set of individual information strategies f consisting of a system of signal distributions $f_t : J \times \Theta \times \mathcal{L} \times \Delta\Theta \times \mathbb{R}^J \to \Delta X$, as well as action strategies σ consisting of a system of mappings $\sigma_t : J \times X \times \Delta\Theta \times \mathcal{L} \times \mathbb{R}^J \to J$, such that:*

- ***Utility maximization**: Agents solve the problem (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 simplifies considerably the characterization of agents' location choices by allowing us 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 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 assimilated as a choice of a distribution of recommendations. Each

[17] This contrasts with the DeGroot heuristic 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 DeGroot model, such an individual would be influential as long as his expectation is different than others'.

[18] See Appendix B.1.1.

signal realization essentially reduces to an instruction about the location to choose.[19]

Let $\omega_t = (\theta_t, L_{t-1}, \pi_t, l_{t-1}, \varepsilon_t)$ denote the state characterizing any specific decision node. A mobility rule p is a system of distributions $p_{l_{t-1}l_t}(\theta_t, L_{t-1}, \pi_t, \varepsilon_t) = \Pr(l_t|\omega_t)$ over J, one for each possible ω_t, interpreted as the probability of moving to l_t at the node ω_t.

Lemma 1. *Any equilibrium strategy (f, σ) solving the dynamic rational inattention problem leading to (10) generates a choice rule p solving:*

$$\max_p \mathbb{E}_{\theta^t,\varepsilon^t}\left[\sum_{t=1}^{\infty}\delta^t\left(u_{l_{t-1}l_t}(\theta_t, L_t, \varepsilon_t) - I_{l_{t-1}}(L_{t-1}, \pi_{l_{t-1}t}, \varepsilon_t)\right)\right], \tag{15}$$

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

$$I_j\left(L_{t-1}, \pi_{jt}, \varepsilon_t\right) = \lambda_j\left(H\left(\pi_{jt}\right) - \sum_k q_{jkt}(L_{t-1}, \pi_{jt}, \varepsilon_t) H\left(\pi_{jkt}\right)\right), \qquad \forall j, \tag{16}$$

and $q_{jkt}(L_{t-1}, \pi_{jt}, \varepsilon_t) = \int_\theta p_{jkt}(\theta, L_{t-1}, \pi_{jt}, \varepsilon_t)\pi_{jt}(\theta)d\theta$ is the ex-ante probability of receiving the recommendation to move to k. Population and beliefs follow the laws of motion (12) and (13) - (14) respectively, with $\bar{p}_{jkt}(\theta_t, L_{t-1}, \pi_{jt}) = \mathbb{E}_{\varepsilon_t}\left[p_{jkt}(\theta_t, L_{t-1}, \pi_{jt}, \varepsilon_t)\right]$.

Proof. See Appendix B.1.2. □

Accordingly, I will henceforth 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 2. *There exists a solution to the dynamic rational inattention equilibrium.*

Proof. See Appendix B.1.3. □

Proposition 2 extends the existence result derived in Steiner et al. (2017) to the case of a non-finite state space with a continuum of agents and endogenous payoffs by ensuring that the strategy space is compact, and that the boundedness of payoffs together with discounting ensure that agents' objective functions are continuous in their strategies.

2.6 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 steady state. A stochastic steady state consists of a mobility rule and beliefs 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 every period as new productivity vectors are realized. But population and beliefs in each location evolve within an ergodic distribution, and we can drop the t subscript for the mobility rule and beliefs.

[19] In dynamic models, we need to make sure that there is no incentive to want a richer signal structure at t than just recommendations, 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 acquired by the agent.

Definition 2. *A stochastic steady state equilibrium is a mobility rule* $p : J \times \Theta \times \mathcal{L} \times \Delta\Theta \times \mathbb{R}^J \to J$*, as well as beliefs* $\pi : J \times \mathcal{L} \times \Delta\Theta \times \mathbb{R}^J \to \Delta\Theta$*, such that* p *is solution to* (15)*, while population and beliefs follow the laws of motion* (12) - (14) *and belong to ergodic distributions.*

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}, making the description of the steady-state behavior intractable. However, if prior beliefs are "close enough", rationally attentive agents make information acquisition decisions that result in the same posterior beliefs. If one agent is more pessimistic than another about the payoffs in some region, she will be less likely to move there, but conditional on moving, the two agents will have the same posterior beliefs. This property of *locally invariant posteriors* was shown in the context of a static model by Caplin and Dean (2015).[20]

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 important 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. I discuss the plausibility of this assumption in Appendix B.6.[21]

Assumption 2. *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.*

We are now ready to characterize the solution to the stochastic steady-state. As shown in Proposition 3, under Assumption 2, prior beliefs only depend on the population distribution, so that the effective states upon which mobility decisions are made reduce to $(\theta_t, L_{t-1}, \varepsilon_t)$. Furthermore, the mobility rule takes a dynamic logit form, with the future value also solving a Bellman equation akin to dynamic logit models. However, the expression contains an endogenous bilateral shifter in the form of the ex-ante moving probability q_{jk}. This shifter plays the role of a *predisposition* towards a move from j to k that reflects the prior beliefs held by agents in j about payoffs in region k. If the expected probability of moving to k is large according to the prior beliefs, then moving to k is more likely, irrespective of the current realization

[20] 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 agents chooses some posterior beliefs, other agents with priors that can also sustain these posterior will choose the same posteriors.

[21] If Assumption 2 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}.

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 region k tends to offer high payoffs for agents in j, either because of high productivity, amenity or low migration costs, this translates into higher average migration flows between j and k and therefore into a higher predisposition for this flow.

Proposition 3. *Under Assumption 2, steady-state beliefs π only depend on the current population distribution L_{t-1}, and the mobility rule p only depends on $(\theta_t, L_{t-1}, \varepsilon_t)$. Conditional on being in j at $t-1$, conditional on the population distribution by previous origin $L_{t-1} = \{L_{jkt-1}\}_{j,k}$ and θ_t, the location decision at t is independent of θ_{t-1} and location decisions anterior to $t-1$. Moreover, for each agent located in j at $t-1$, the optimal mobility rule $p_{jk}(\theta_t, L_{t-1}, \varepsilon_t)$ can be expressed as:*

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

where $q_{jk}(L_{t-1}, \varepsilon_t) = \int_\theta p_{jk}(\theta, L_{t-1}, \varepsilon_t) \pi_{jt}(\theta | L_{t-1}) d\theta$ 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

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

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

Proof. See Appendix B.1.4. □

In their analysis of migration decisions Kennan and Walker (2011) estimated moving costs that were surprisingly large. As will become clear in the section 4, when agents can acquire costly information about wages at other locations, estimates of the moving costs will decrease: since moves are relatively rare, the predisposition term creates a virtual cost of moving, which the cost identified by Kennan and Walker (2011) combines with the true moving cost κ_{jk}. In addition, Kennan and Walker (2011) considered a counterfactual policy experiment involving a subsidy for moving. With information frictions, the effect of the subsidy will be larger: not only does the subsidy have a direct effect on payoffs, it also increases the predisposition toward moving, thereby lowering the associated virtual cost; the information acquisition induced by the subsidy reinforces the increase in migration.

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}), we need to aggregate these individual decisions by integrating over preference shocks ε_t. Moreover, note that even though we reduced the dependence of equilibrium beliefs to the population distribution only, their actual distribution is undetermined. To make progress, we look for equilibrium beliefs that belong to some particular class of distributions, the Cardell class, that will deliver useful aggregation results. In addition, we assume that preference shocks are also drawn from this class of distributions.[22] The main property of the Cardell distribution is that if a random variable X is drawn from a Type I extreme value distribution

[22] Brown and Jeon (2019) also imposed that beliefs and preferences follow Cardell distributions in their static model of optimal health insurance choice. Dasgupta and Mondria (2018) imposed that productivity shocks follow Cardell distributions in their model of international trade.

and another random variable Y is drawn from a Cardell distribution with dispersion ν, then $Y + \nu X$ is a random variable distributed as Type I extreme value.[23]

Assumption 3. *Equilibrium marginal beliefs $\pi_j(\theta_{kt}|L_{t-1})$ can be approximated by independent Cardell distributions with mean $\mu_{jk}(L_{t-1})$ and same variance $\sigma_j(L_{t-1})^2$.*

Assumption 4. *Preference shocks ε_{ikt} are drawn from independent Cardell distributions with mean zero and dispersion ν.*

Although Assumption 4 is similar to the usual assumption that preference shocks are drawn from Type I extreme value distributions, Assumption 3 admittedly imposes restrictions on the equilibrium behavior of agents by constraining beliefs to belong to a particular distribution and be independent across regions. In Appendix B.7, I relax the assumption of independent beliefs and look for Gaussian beliefs in the case with no preference heterogeneity and a limited number of regions and show that the equilibrium mobility patterns are indistinguishable from the ones obtained under Assumption 3. Before presenting the main result, I denote by $\bar{u}_{jk}$ the average utility flow $\mathbb{E}_\varepsilon [u_{ijk}]$ 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}):

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

Proposition 4. *Under Assumptions 3 and 4, the average mobility rule in the presence of information frictions and preference heterogeneity writes:*

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

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:

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

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

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

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

When the dispersion of preferences ν tends to zero, the solution writes:

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

[23]The relation between the Cardell distribution and the Gumbel distribution is displayed in Appendix E. Qualitatively, the two distributions are very similar.

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

Proof. See Appendix B.1.5. □

The role of information frictions λ_j in altering the responsiveness of migration to local shocks is clearly apparent in (19). The elasticity of migration with respect to observed components of $\bar{u}_{jk}$ is $1/\phi_{jt}$, whereas the elasticity of migration with respect to θ_t is smaller and equal to $(1 - \eta_{jt})/\phi_{jt}$. Since $\eta_{jt} \in (0,1)$, 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 choose to only imperfectly observe it. When information costs λ_j become very large, $\eta_{jt} \to 1$ and the elasticity of migration with respect to θ_t is zero. Conversely, when information costs are zero, beliefs about θ_t become perfectly accurate and $\mu_{jk}(\theta_{kt})$ equals θ_{kt}. As shown by (21), this implies that the elasticity of migration with respect to θ_t is the same as the elasticity with respect to other observed payoffs and 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 payoffs. In Section 4.2, I will propose a strategy to estimate the elasticities $(1/\phi_{jt}, (1-\eta_{jt})/\phi_{jt})$ and 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 is materialized in (19) by the role of mean beliefs $\mu_{jk}(\theta_{kt})$ in shifting the level of bilateral migration flows. Destinations k about which agents in j hold optimistic beliefs will be favored over other regions, creating the same endogenous predispositions towards some regions as was discussed in (17). For example, agents in region j hold optimistic beliefs about region j, which reduces their likelihood of moving to another region. As I will discuss in Section (4.4), omitting to account for information frictions will lead to an overestimation of the actual bilateral migration costs κ_{jk}. The dependence of mobility decisions on mean beliefs will also have implications for mobility patterns. As I will argue in Section 5, the covariance of mean beliefs with the realized productivity varies across pairs of regions and delivers migration patterns that are in line with observed flows. In particular, agents' beliefs will be less responsive to productivity shocks in distant regions, both because they are less likely to interact with well-informed people from these regions, and because they will individually gather less information about them. This will lead to a decreasing migration elasticity with distance.

3 Data Sources

In this section, I present the main sources of data and provide descriptive statistics on the migration patterns in Brazil. I provide more information on the construction of the sample and descriptive statistics in Appendix A.

The main source of data is the Relação Anual de Informações Sociais (RAIS), which is collected annually by Brazilian Ministry of Labor and contains matched employer-employee information for every formally employed worker in Brazil. It includes demographic, occupational and income characteristics of employees, with individual identifiers so that workers can be followed over years. I use 16 consecutive year of data, corresponding to the period between 2000 and 2014. It 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 at which the worker is employed.[24] Specifically, I use information on the micro and mesoregion in which the plant is located.[25]

It is not infrequent that workers in the sample 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 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.

The data contains no information on the location of 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 of potential workers of whether to enter the labor force or acquire college education. For this reason, I limit the data to observations that correspond to workers that are over 25 years of age, since, for the majority of the population, educational decisions are taken before this age. Similarly, I do not model the retirement decision of workers and, consequently, I limit the data to observations that correspond to workers that are 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 a measure of population by mesoregion by aggregating the population of all municipalities included in the corresponding mesoregion.

Finally, I collect information on the degree of internet penetration by microregion. Specifically, from the *Agência Nacional de Telecomunicações*, ANATEL, a government agency in charge of regulating and supervising telecommunications in Brazil, I obtain information on the number of broadband connections by municipality and year 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 use again the municipality-level population data to construct an equivalent variable at the mesoregion level as a population-weighted average of the per capita number of broadband connections in the municipalities included in the corresponding mesoregion. 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 and year prior to 2007.

The resulting dataset includes 372,454,979 worker-year pairs that correspond to 45,958,805 workers. According to the RAIS data, the 137 mesoregions in which the geography of Brazil is divided had in 2014 on average close to 251,000 legal workers (the median microregion had close to 62,000 legal workers). The most populated mesoregions are located in the South and along the coast. The average yearly rate of

[24] 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.

[25] 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 themselves grouped into 137 mesoregions.

migration across mesoregions is 3.4%.

4 Estimation of the Model

In this section, I structurally estimate the model, guided by the mobility rule derived from the previous section. I estimate the model separately over two time periods: 2000-2007 and 2008-2014. For each of these periods, I follow a two-step procedure: first, I estimate the production parameters that can be inferred directly from the data, and are separate from the rest of the system; 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. I find that information costs are higher in less developed regions and have decreased between the two periods. I show that allowing for information frictions rather than assuming complete information leads to bilateral migration costs that are 40% smaller and a 20% lower dispersion of preference heterogeneity.

The structural estimation uses data from RAIS described in Section 3. From individual records or earnings and location every year, I construct yearly bilateral migration flows between each of the 137 mesoregions in Brazil, and average wages in each region and year.

4.1 Step 1: Production parameters

I begin by estimating $J+3$ parameters that can be obtained directly from the data by relying on assumptions with respect to the production side of the model only. These are 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 to equal the national level labor share computed from national statistics over each period.[26]. The inverse labor share α in Brazil is stable between 2000 and 2014 at a value of 0.40. Since the labor share governs the extent to which payoffs depend on local population, it plays an important role in the estimation of information frictions and preference heterogeneity. For this reason, I estimate the model parameters for alternative values of α over a range of plausible values and discuss the implications for the results in Appendix B.8.

I then turn to the estimation of the productivity process. The Brazilian economy is growing over the sample period. The total population recorded in RAIS is also growing every year, both from demographic change and increased transitions to the formal labor market. To be able to interpret the data as closely as possible to a steady state with constant total population, I project wages on year fixed effects and normalize all population stocks so that the total population in the economy is constant to its 2000 level over the sample. I can then estimate the AR(1) process associated with the observed wages cured from year fixed effects and corrected for population $y_{kt} = \log w_{kt} + \alpha \log L_{kt}$. From (3) and (6), the corrected wages can be expressed as:

$$y_{kt} \equiv \log w_{kt} + \alpha \log L_{kt} = A_k + \theta_{kt}. \tag{23}$$

The baseline productivities are therefore recovered as the average wage of y_{kt} over the period, since $\mathbb{E}[\theta_{kt}] = 0$. The persistence of the productivity process ρ is equal to the covariance of y_{kt} and y_{kt-1} divided

[26]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.

by the variance of y_{kt}. The volatility of the productivity process is then computed as $(1-\rho^2)\text{Var}(y_{kt})$. Finally, I recover the productivity shocks θ_{kt} implied from the observed y_{kt} and estimated productivities A_k as the residuals from (23). 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 and 0.69 in the first and second periods. The volatility of the productivity process σ_ξ^2 is 0.33 and 0.38 in the first and second periods.

4.2 Step 2: Simulated Method of Moments

In the second stage, I use the method of simulated moments to estimate the remaining parameters $\omega = (\nu, \lambda_j, B_j, \kappa_{jk})$ which consist of the dispersion of preferences, the information costs, amenities and migration costs. The total number of parameters to be estimated in this step is large: there are as much as $2J + (J-1)^2 = 18770$ of them.[27] A grid search over the parameter space is therefore not computationally possible in this context. 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 derived 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 basis of the 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 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 (4), we get the following expression:

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

where $D_{jk} = B_k - B_j - \kappa_{jk}$ is a composite bilateral "resistance" term combining the contributions of amenity differences between regions as well as migration cost on migration flows. The error term $e_{jkt} = \left(\zeta_{jjt} - \zeta_{jkt}\right)/\phi_{jt}$ is composed of expectational errors about future values $\zeta_{jkt} = \delta\left(V_{kt+1} - E_{j,t}V_{kt+1}\right)$. Equation (24) expresses the "gravity" structure of migration patterns in the sense that the magnitude of bilateral flows is increasing in the wage – and future value – differentials between any two regions, and 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 expectational errors difference.

The regression equation (24) serves as the basis for the estimation by simulated method of moments. To make this apparent, denote the outcome variable by $y = \log\left(\bar{p}_{jkt}/\bar{p}_{jjt}\right)$, the set of coefficients by $\beta = (\eta_{jt}/\phi_{jt}, 1/\phi_{jt}, \{D_{jk}/\phi_{jt}\})$, the set of regressors by $X = (\Delta\chi_{jkt}, \Delta w_{jkt} + \delta\Delta V_{jkt}, \{\mathbb{1}_{jk}\})$, where $\Delta x_{jkt} = x_{kt} - x_{jt}$ and $\mathbb{1}_{lm,jkt}$ is a dummy variable equal to 1 if $(j,k) = (l,m)$. With this notation, (24) rewrites:

[27] 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.

$$y = X\beta + e. \tag{25}$$

Since this error term e only arises because of the irreducible uncertainty about future productivity θ_{t+1}, it is orthogonal to all other regressors. As a result, the coefficients $\hat{\beta}$ obtained from the estimation of (25) by ordinary least squares (OLS) are known to be the method of moments estimator for the moment condition:

$$\mathbb{E}\left[X\left(y - X\beta\right)\right] = 0. \tag{26}$$

Note that there is a direct mapping between the coefficients of the regression for each time period t and the parameters of interest $\omega = (\nu, \lambda_j, B_j, \kappa_{jk})$. According to the model, the coefficients ϕ_{jt} and η_{jt} should vary over time because they depend on the variance of agents' beliefs at t, which itself depends on the population distribution. However, allowing for time varying region specific coefficients in (24) is very demanding on the data. Since in practice the variance of beliefs is not very sensitive to the local population – the mean of beliefs is the main driver of the variation of beliefs with respect to the local population – I estimate only the average η_j and ϕ_j for each region. With this simplification, it is easy to show that there exists a unique mapping from estimates $\beta = (\eta_j/\phi_j, 1/\phi_j, \{D_{jk}/\phi_j\})$ to ω.[28]

We can now see that a satisfactory set of parameters ω would be such that the belief functions $\mu_{jk}(L_{t-1})$, value functions $V_j(\theta_t, L_{t-1})$ arising from the model with parameters ω, would be such that when we estimate the equation (24) by OLS with the migration flows and wages observed in the data and the beliefs and values evaluated at the productivity and population observed in the data, the estimated coefficients $\hat{\beta}$ would map exactly to $\hat{\omega}(\hat{\beta}) = \omega$. This is exactly the logic behind the iterative algorithm that I employ to recover ω. First, I guess initial values $\omega^0 = (\nu^{(0)}, \lambda_j^{(0)}, D_{jk}^{(0)})$ for the parameters to be estimated.[29] Second, I simulate the model using the production parameters $(\alpha, \rho, \sigma_\xi^2)$ estimated in the first block, and the current parameters $\omega^{(n)}$. This delivers mean beliefs functions $\mu^{(n)}(\cdot)$ and values $V^{(n)}(\cdot)$. Third, I evaluate beliefs and values at observed populations L_{t-1}^{obs} and recovered θ_t^{obs}, and estimate (24) by OLS.[30] Fourth, from the estimated η_j and ϕ_j and fixed effects, I update the parameters to $\omega^{(n+1)}$ and return to the second step until convergence to ω^*. By construction, for ω^*, the method of moments objective function associated to the moment conditions (26) is minimized. I describe the iterative algorithm in more detail in Appendix C.1. Despite theoretical results on the convergence of this algorithm, I find that in practice, the algorithm converges quickly and always to the same solution ω^*.

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 and 2.62 in the first and second period. Interestingly, the dispersion in preferences appears to have increased over time while information costs decreased. This reflects the fact that migration flows have become slightly less directed towards larger 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 amenity to zero, and

[28] See Appendix C.1 for more details.

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

[30] 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 parameter values of the productivity process parameters, regional baseline productivities, amenities, and information costs, as well as bilateral migration costs.

obtain that the standard deviation is 0.83 and 0.91 for the first and second period. The average migration costs are 0.56 and 0.47 in the first and second period. The slight decline in migration costs between the two periods is consistent with the overall increase in migration flows over the time period.

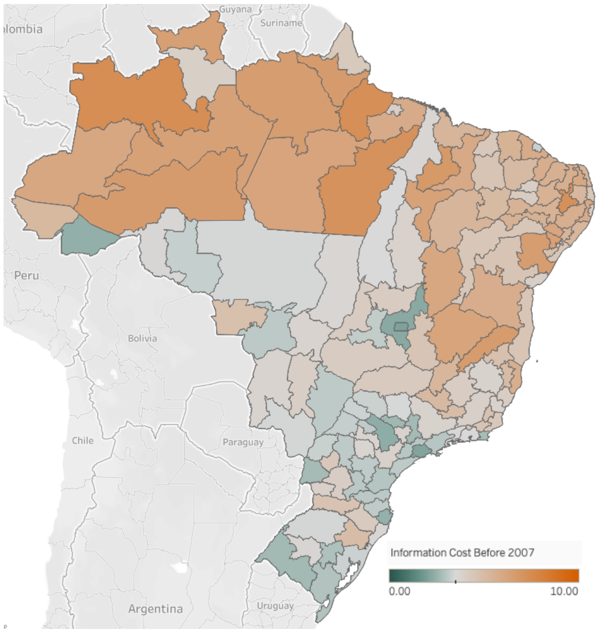
4.3 Information Costs and Internet Access

I find that information costs are on average equal to 3.11 and 2.23 in the first and second period respectively. The standard deviation in the information costs is 1.43 and 1.32 in the first and second period. Figure 1a and Figure 1b display the estimated information costs λ_j for each of the 137 mesoregions, for the first and second period. The information costs appear strongly correlated with the economic development of regions, with the lowest costs of information access obtained for 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 much higher cost of information. It is worth emphasizing that these information costs were recovered as fixed-slopes 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 migration flows from these regions appear to be relatively more responsive to the unobserved productivity shocks than the other regions.

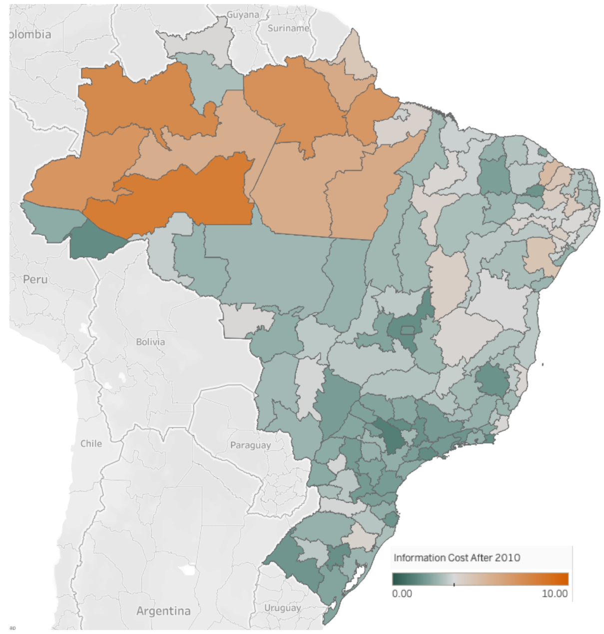
Since these information costs are novel parameters that have not yet been estimated in the literature, it is of interest to describe their variation along observable regional characteristics. From inspecting Figure 1a and Figure 1b, these information costs appear to be potentially correlated with the population density and average income of the regions. 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 the geographic accessibility to the region. Motivated by these remarks, I propose a simple parameterization of information costs:

$$\lambda_{jt} = \ell_1 int_{jt} + \ell_2 \log w_{jt} + \ell_3 \log popdens_{jt} + \varsigma_j + u_{jt}, \tag{27}$$

where $t \in \{1, 2\}$ corresponds to each of the two periods 2000-2007 and 2008-2014, int_{jt} is the average



(a) Information Costs 2000-2007



(b) Information Costs 2008-2014

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 (27) in first differences by Ordinary Least Squares (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 will be interested in reproducing the plausible decrease in information costs brought about by the expansion of internet access at the local level. From inspecting (27), one may suspect that there could be unobserved time varying factors in the error u_{jt} that would be 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 information cost, and facilitated internet expansion. We may expect that these omitted variables would lead to an upward bias in the estimation of the causal effect of internet penetration on the information cost.

To try and 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.[31]

[31]"Backbones" are national trunk infrastructure that brings traffic from international submarine cables in coastal regions

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.

Proximity to these key elements of the internet infrastructure is essential to provide a high quality of broadband connection. 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 us with 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.3. I report the results from the instrumental variable regression in column 2 of Table 2. As expected, 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 the model for the magnitude of the migration costs and preference heterogeneity. Since information frictions are a source of both endogenous migration costs and limited migration elasticity, we expect that 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 information frictions are not present, so that $\lambda_j = 0$ 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 the use of renewal actions and the comparison of 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 that the difference in the payoffs they offer can be expressed as a function of the wages in the visited regions at $t+1$. If we 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, we can write the following estimation equation:

$$\Lambda_{jkt} = \frac{1}{\nu} \left(\Delta \log w_{jkt} + D_{jk} + e_{jkt} \right), \tag{28}$$

to inland parts of the country. Backbones consist of high-capacity fiber optic cables.

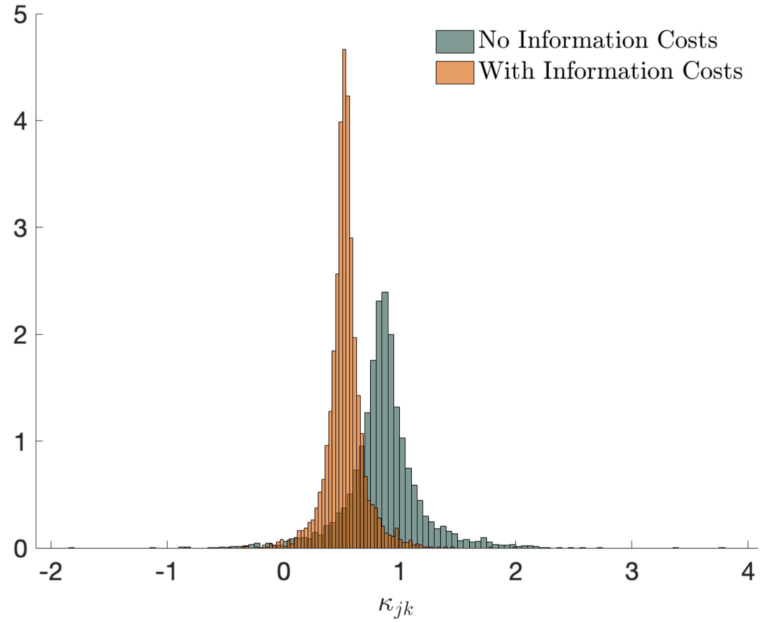


Figure 2: Estimated Migration Costs with and without Information Cost

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 which 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. This is described in Appendix C.2. I obtain almost exactly the same estimates as in the direct estimation, which confirms the accuracy of the iterative estimation method.

The distribution of bilateral migration costs resulting from the estimation of (28) – after projecting on origin and destination fixed effects to net out amenities – is displayed in Figure 2 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 $10,382$ of them have positive flows recorded more than once over the 7 years of the first period. This implies that no fixed bilateral resistance term is estimated for the pairs with no consistent flows. In figure 2, the average migration costs among the $10,382$ estimated bilateral costs is 0.94. In comparison, I report the distribution of the estimated migration costs for the same $10,382$ origin-destination pairs in the model with information frictions. As mentioned in Section 4.2, the average migration cost in the first period is 0.56, so that the average migration cost is 40% smaller once we allow for information frictions. This illustrates the quantitative relevance of the endogenous predispositions emanating from the information frictions.

To convert the magnitude 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 have to be such that $\log(w + \Delta w) =$

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

$\log w - 0.56$, 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 (28) in the first period is equal to $\nu = 2.92$, to be compared with 2.31 in the model with information frictions. The preference heterogeneity is then 21% smaller once we recognize the role of information frictions. For the second period, I estimate $\nu = 3.12$, to be 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

Migration costs have been estimated 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:

$$\kappa_{jk} = \beta_1 \log dist_{jk} + \beta_2 \log traveltime_{jk} + \beta_3 contiguous_{jk} + e_{jk}. \tag{29}$$

Table 3 reports the results from estimating (29) in both periods. Migration costs are increasing with travel time and distance, and are smaller between contiguous regions. There is a slight decline in the magnitude of the coefficients between the two periods, indicating migration costs appear to have become slightly less dependent on distance and travel time. In 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 distinct exercises that illustrate the success of the model at describing migration patterns relative to a model with complete information. In the first exercise, I uncover a set of new facts on migration patterns in Brazil, namely that the migration elasticity with respect to wages is decreasing with the distance between regions, increasing with the intensity of the past migration flows connecting them, and increasing with the internet penetration at the origin. I replicate the same empirical exercise in the 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 features delay that is longer for more distant origins and origins with lower internet penetration. I replicate a number of actual positive local shocks in the model and show that only the model with information frictions is able to reproduce this differential delay. Taken together, these two exercises offer results support of both the qualitative and quantitative relevance of the new mechanisms introduced in the model.

Before describing in detail 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 the observed migration patterns. This is in large part thanks to their flexibility with respect to the bilateral migration costs κ_{jk}, which allow to match exactly the average levels of bilateral migration flows between any two regions. Hence, in order to truly 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 to recover the migration elasticity $1/\nu$, described in (28). We can however allow for some heterogeneity in the migration elasticity along some variable Z_{jkt}:

$$\Lambda_{jkt} = \beta_1 \Delta \log w_{jkt} + \beta_2 Z_{jkt} \times \Delta \log w_{jkt} + \beta_3 Z_{jkt} + D_{jk} + e_{jkt}. \tag{30}$$

For any variable Z_{jkt}, the model with complete information would predict that $\beta_2 = \beta_3 = 0$. However in the presence of information frictions, from the estimation 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.

First, in the presence of information frictions, we can expect the migration elasticity to 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 decide to 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 $t-1$, 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 becomes more optimistic. Since the productivity in k is persistent, it is likely to be high also at t, and agents in j are now likely to think it is 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 in turn make migration more responsive to opportunities in k. In the 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. As discussed above, 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, we expect people with access to internet to be able to gather more accurate information about migration opportunities. This should induce them to be more likely to take advantage of positive shocks and 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 the 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 the odd columns of Table 4, I report the results from estimating equation (30) over the second 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, the log number of individuals who moved from k to j in the past year. In the even columns, I report the results from estimating equation (30) with migration flows and wages predicted by the model using parameters associated with the second period 2008-2014. Once I solved 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. For the sake of conciseness, I do not report in Table 4 the results from the estimation of (30) with the model with no information frictions. As expected, the coefficients on the interactions from these regressions are all precisely estimated zeros and are reported in Appendix D.2.

In column 1, the estimated coefficient of 0.391 in front of the income gap corresponds to the migration elasticity between pairs that are adjacent, while the negative coefficient in front of the interaction of the

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

	(1) Data	(2) Model	(3) Data	(4) Model	(5) Data	(6) Model
Income gap	0.391^a	0.342^a	0.281^a	0.306^a	0.362^a	0.391^a
	(0.012)	(0.007)	(0.016)	(0.008)	(0.018)	(0.008)
Log dist × Inc. gap	-0.012^a	-0.009^a				
	(0.004)	(0.002)				
Internet × Inc. gap			0.132^a	0.101^a		
			(0.015)	(0.004)		
Log Past Flows × Inc. gap					0.0031^a	0.0063^a
					(0.0013)	(0.0007)
Observations	68,152	68,152	68,152	68,152	68,152	68,152
R^2	0.179	0.815	0.157	0.813	0.168	0.882

a denotes 1% significance, *b* denotes 5% significance, *c* denotes 10% significance. In parenthesis, 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. In each regression, origin-destination and year fixed effects are included. The income gap is the different 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.

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 distant by the average distance of 500km, the estimated elasticity is therefore 0.3164. Note that it 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. I 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 strongly emphasizes the quantitative relevance of information frictions, as it indicates that most of the decline in the responsiveness of migration flows with distance can be accounted for by the less effective acquisition and transmission of information between remote regions.

In column 3, 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 4, 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 give rise to this positive interaction coefficient with internet access is another success of the model, especially considering the fact that the local information costs λ_j were estimated as fixed effects without imposing any relationship with respect to internet access.

In column 5, 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 the data, the 25^{th} percentile across all pairs for the fraction of the population from a given origin is 3.10^{-8}, while the 75^{th} percentile is 7.10^{-4}, leading respectively to migration elasticities of 0.308 and 0.340. In column 6, I report the results from the same regression in the model. As was expected, the migration elasticity with respect to wages in some destination is significantly larger when this destination is connected to the origin via large recent migration flows. The effects are about twice larger in the model than in the data, which is not surprising since the past migration flows is exactly the metric that governs the adjustment of beliefs through local information sharing in the model. Taken together, these results show that the introduction of information frictions in the model delivers predictions that match important characteristics of migration flows.

5.2 Delay in 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 origin 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 a function almost exclusively of their migration shares between their origin and the location of the shock prior to 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 are susceptible to 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 t, I define formally 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 $p_{jkt+s}^{(\underline{t},\bar{t})}$ from j to k at $t+s$ with respect to a horizon $(\underline{t}, \bar{t})$ writes:

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

I also define the expected time of migration between j and k over a horizon $(\underline{t}, \bar{t})$ as $\sum_{s=0}^{\bar{t}} sp_{jkt+s}^{(\underline{t},\bar{t})}$. I then consider 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 j' if of all the migrants who will move to k between t and $t + \bar{t}$, the migrants from j tend to move earlier than the ones from j'.

For any 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:

$$p_{jkt+s}^{(\underline{t},\bar{t})} = \sum_{s'=-2}^{4} \mathbb{1}\{s = s'\} \left(\gamma_{1s'} int_{jt} + \gamma_{2s'} \log dist_{jk} + \gamma_{3s'} \log p_{jk}^{past}\right) + u_{jkt+s} \tag{31}$$

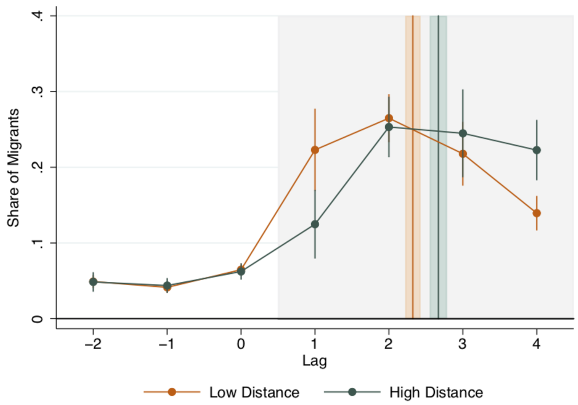
where t indicates the year in which the labor demand shock in the region of interest took place, int_{jt} is the internet penetration in region j at the year of the shock, $dist_{jk}$ is the distance between j and k, p_{jk}^{past} is the average migration probability from j to k in the three years preceding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

Figure 3a and 3b display the estimates of $\{\gamma_{1t+s}, \gamma_{2t+s}\}$ obtained when focusing on the delay with which migrants reacted to a particular local shock. 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% percentile, labeled as "Low", or to its 75% 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 dark thin vertical lines indicate the expected time of migration between j and k once the shock is realized, and the light-colored thick vertical lines illustrate the corresponding 95% confidence intervals. Everything else equal, workers living in geographically close regions or in regions with higher broadband internet penetration tend to react faster to the positive labor demand shocks happening in Ipojuca.

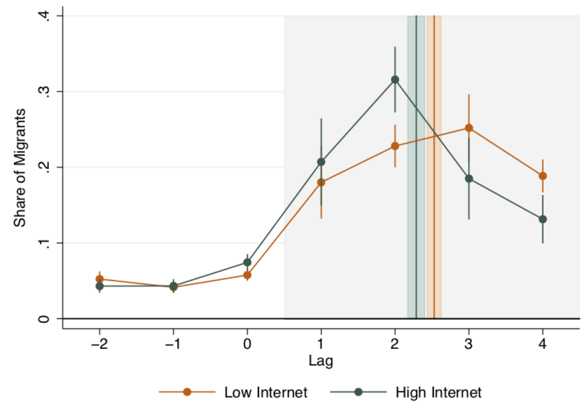
I then run the regression (31) in the model, using first the model with information frictions. Selecting the destination of interest k to be the mesoregion of Recife, I evaluate the migration flows predicted by the model under the observed population and recovered productivity vectors. The recovered productivity features a persistent increase starting in 2009 (see Appendix 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 3c illustrates that the estimated coefficients on distance lead to a significant delay for regions that are more distant to Recife. Figure 3d shows that the model also predicts a significant delay from regions with lower internet access.

In the model with no information frictions, Figures 3e and 3f reveal that the migration patterns resulting from the exact same simulated shock in do not result in any significant delay along distance or internet access once we control for the past migration shares. This result confirms the success of the model

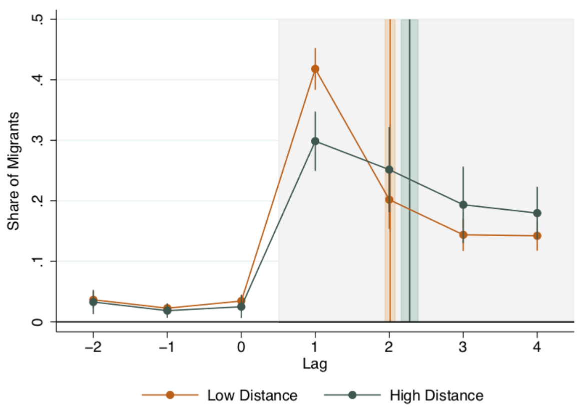
Figure 3: Delay in the Migration Response to the Local Shock in Ipojuca



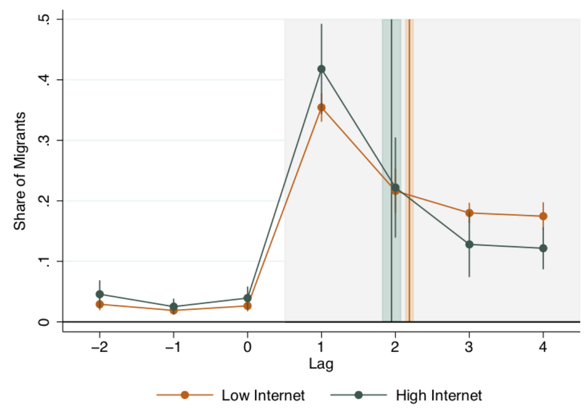
(a) Distance: Data



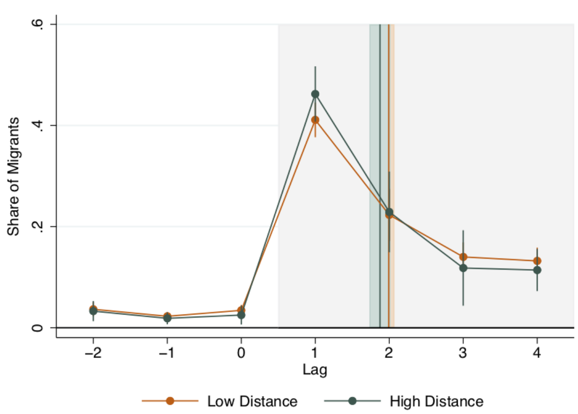
(b) Internet: Data



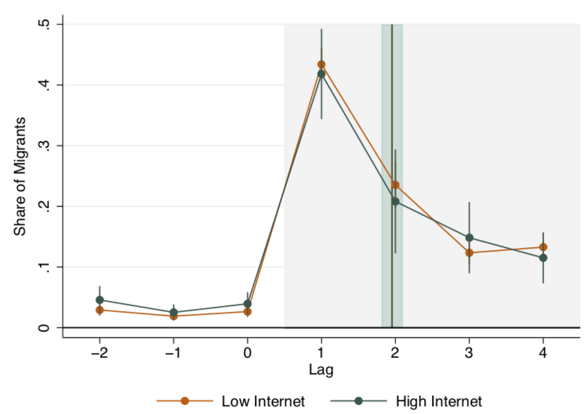
(c) Distance: Model with information frictions



(d) Internet: Model with information frictions



(e) Distance: Model without information frictions



(f) Internet: Model without information frictions

Note: The standard errors of the estimates of the implied expected probabilities and the expected expected time of migration are computed using the Delta method and standard errors for the estimates of the parameter $\{\gamma_{1t+s}, \gamma_{2t+s}\}$ clustered by year and mesoregion of origin.

with information frictions in providing a rich and accurate description of the migration patterns observed in the data, here in the case of 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 over the period.

6 Counterfactual Exercises

I have discussed in detail several implications of allowing information frictions to affect migration decisions. The main ones 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 shown that these predictions are verified in the data, and that the magnitude of the effects in the estimated model are close to the observed ones. To illustrate further the quantitative implications of the model for the spatial allocation of workers and welfare, I conduct in this section a number of counterfactual exercises. First, I compare the outcomes in the economy estimated in the first period to a hypothetical one in which the information frictions would be reduced to zero. Second, instead of reducing the information costs from their estimated level in the first period to zero, I reduce them by the amount corresponding to the estimated effect of increased internet access in the second period. For each of the two 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 to be expected if information frictions were to be completely removed. This exercise will help us get closer to answering 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? Only now that we have a model at hand in which information frictions have been given a precise role can we consider altering the intensity of these frictions.

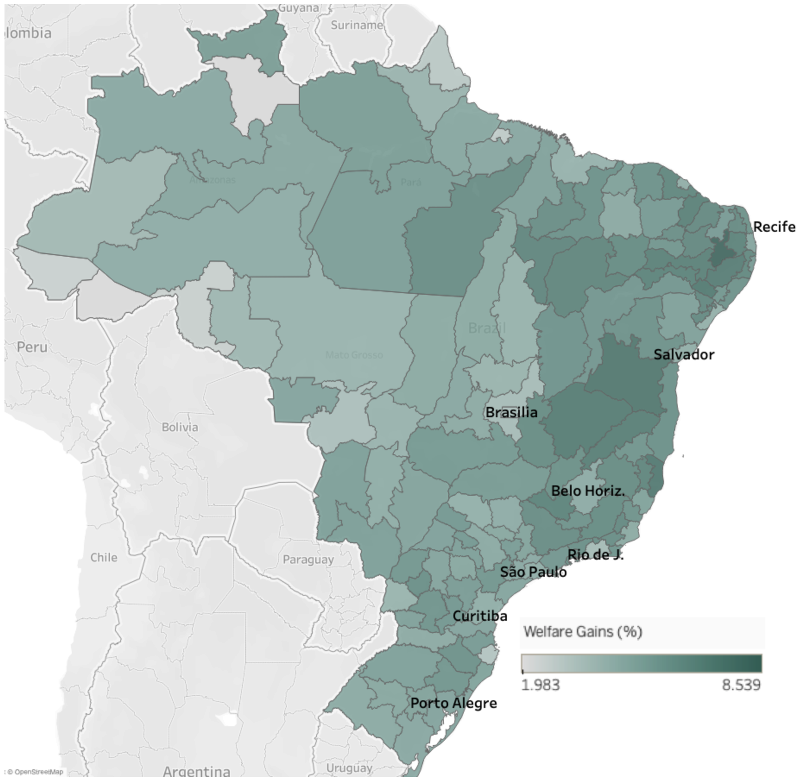
To evaluate the effects of removing all information frictions, I first compute the expected value of being in a given region in the economy corresponding to the steady state of the first period 2000-2007. In the stochastic steady state, the value of residing in a given region varies over time as different productivity shocks and population vectors are realized. Therefore, I report the expected value over all possible realizations of (θ_t, L_{t-1}) of the ergodic distribution. Denoting by $\mathcal{V}_j$ the expected value of a region, we can write:

$$\mathcal{V}_j = \mathbb{E}_{\theta_t, L_{t-1}} \left[\sum_k \bar{p}_{jk}(\theta_t, L_{t-1}) \left(\bar{u}_{jk}(\theta_t, L_{t-1}) + \delta \bar{V}_k(\theta_t, L_{t-1}) \right) - I_j(\pi_j(L_{t-1})) \right]. \tag{32}$$

The expected value $\mathcal{V}_j$ can be computed simply by simulating the model over a very 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 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 are set to zero. In this new stochastic steady state, I compute the expected values $\mathcal{V}'_j$ in each region.

Figure 4 displays the percentage change in the expected value of each region $\Delta\mathcal{V}_j/\mathcal{V}_j$ between the two

Figure 4: Welfare Gains by Mesoregion after Setting All Information Costs to Zero



equilibria, where I now define $\Delta X = X' - X$. The average welfare gains across regions are large and equal to 5.55%. There is important heterogeneity across regions, with gains ranging from 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 1a, 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 the regions with initially quite low information costs, such as Brasilia, São Paulo and Rio de Janeiro. This is mostly verified, although some regions in the Amazon do not seem to benefit as much as expected given their large initial information costs.

In the steady state with complete information, the standard deviation of earnings across space decreases by 7%. 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.

Interestingly, the overall migration flows, computed as the fraction of the population moving to another region every period, decrease by 4% in the model with complete information. This net decline in migration flows masks two countervailing forces, both linked to the reduction in the rate of mistakes made by agents. First, and working against 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 effect was a reason for agents to stay in their current region by default. Second, agents now only move to regions that offer high payoffs, and no longer visit a region by mistake so that they would have to move again soon later. The persistence in the productivity shocks allow workers to benefit from a mobility decisions for several periods, reducing their propensity to move often. 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 denoting by $\bar{p}_j = (\bar{p}_{jk})_{k=1,\ldots,J}$ the vector composed of the mobility probabilities for all destinations and states, we can express the difference in expected value in a region j from (32) as:

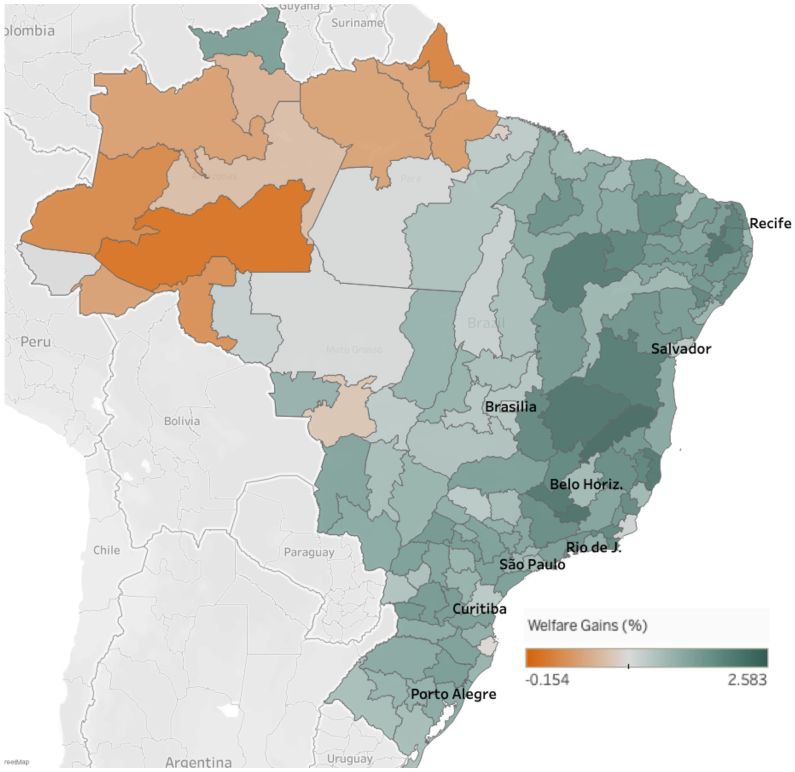
$$\Delta\mathcal{V}_j = \underbrace{\mathbb{E}_{\theta,L}\left[\Delta\bar{p}_j \cdot \bar{U}'_j\right]}_{\text{better sorting}} + \underbrace{\mathbb{E}_{\theta,L}\left[\bar{p}_j \cdot \Delta\bar{U}_j\right]}_{\text{better outcomes}} - \underbrace{\mathbb{E}_L\left[\Delta I_j\right]}_{\text{lower info cost}} . \tag{33}$$

The first term in the decomposition can be understood as the welfare gains arising from the potentially 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 in comparison to the initial equilibrium, maintaining the payoffs at their new value $\bar{U}'_j$. This better sorting can come about thanks to the improvement in the precision of information held by agents, 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 procured by the decline in the utility spent for acquiring information. This decline can arise for two reasons. The first, direct, is simply due to the lower information cost λ_j, allowing workers to reduce their uncertainty about payoffs at a lower cost. The second 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 may 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 the "better outcomes" and 21% to the "lower information costs". It is quite remarkable that most of the gains seem to be occurring through 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 turn to the evaluation of a counterfactual decrease in information costs in each region by a magnitude equal to the estimated contribution of increased 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 amount equal to the local internet access in 2014 (assuming that the internet penetration was zero in the years 2000-2007), multiplied by the estimated coefficient ℓ_1

Figure 5: Welfare Gains by Mesoregion after Expansion of Internet Access



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 is conducted under the assumption 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 with a parallel expansion in access for firms, and that local firms' performance may have been altered by the use of internet services. There is an extensive literature studying the effects of the development of information and communication technologies (ICT) on the organization of production. 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 one resulting from the expansion of internet access, but without affecting the local productivity process.

Figure 5 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 2.01%, 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 expected value of 0.15%, despite experiencing an increase in internet penetration of 23 percentage points.

To understand why some regions may not benefit from the episode of internet expansion, it is useful to decompose their welfare gains into the three sources 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 utility to gather information than they did before. This happens because the information they can obtain from their local network has deteriorated. Indeed, in the new equilibrium, the population of Manaus has decreased by 7%, reflecting the fact that fewer workers from other regions now decide to locate there. These 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 were more likely to have made a "mistake" by moving there. With the supply of well informed visitors impoverished, the 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 available.

One additional reason for which outcomes do not improve enough in Manaus to compensate 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 which the structure of information about opportunities in other regions is determined in equilibrium along with earnings and mobility patterns. I have argued that it is possible to model in a tractable way both the incentives that agents have to acquire information about some regions using the rational inattention framework, and the possibility for workers to benefit from the information circulating within their local networks using a simple social learning rule. When agents are faced with costs of acquiring information, they tend to limit their scope and become less likely to move to regions offering lower expected payoffs. Accounting for these endogenous default rules that hinder mobility appears to be quantitatively important for the estimation of bilateral migration costs. It leads to significantly lower estimates of these costs and helps explain why the previous literature has obtained implausibly large migration costs.

I show that the implications of the 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 if local networks can provide little relevant information, will be less likely to respond.

I discuss how policies that could reduce information frictions by a plausible amount have the scope for generating 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 appear far from trivial, with some regions risking to become “information traps” in which agents struggle to gather accurate information.

The role of information frictions in migration decision is likely to be more complex than the model I put forward in this paper and could be contrasted more fully with the data. 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.

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. and David Dorn, "The Growth of Low-Skill Service Jobs and the Polarization of the US Labor Market," *American Economic Review*, August 2013, *103* (5), 1553–97.

_ , _ , and Gordon H. Hanson, "The China Syndrome: Local Labor Market Effects of Import Competition in the United States," *American Economic Review*, 2013, *103* (6), 2121–2168.

Bah, Tijan and Catia Batista, "Understanding the Willingness to Migrate Illegally: Evidence from a Lab in the Field Experiment," *NOVAFRICA Working Paper*, 2018, *1803.*

Bartik, 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.

Beaudry, Paul, Mark Doms, and Ethan Lewis, "Should the Personal Computer be Considered a Technological Revolution? Evidence from U.S. Metropolitan Areas," *Journal of Political Economy*, 2010, *118* (5), 988–1036.

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

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.

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.

Cardell, N. Scott, "Variance Components Structures for the Extreme-Value and Logistic Distributions with Application to Models of Heterogeneity," *Econometric Theory*, 1997, *13* (2), 185–213.

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, 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.

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

Feyrer, James, Erin T. Mansur, and Bruce Sacerdote, "Geographic Dispersion of Economic Shocks: Evidence from the Fracking Revolution," *American Economic Review*, 2017, *107* (4), 1313–1334.

Fosgerau, Mogens, Emerson Melo, Andre de Palma, and Matthew Shum, "Discrete Choice and Rational Inattention: A General Equivalence Result," *Technical Report*, May 2019.

Fujiwara, Thomas, Eduardo Morales, and Charly Porcher, "A Revealed-Preference Approach to Measuring Information Frictions in Migration Decisions," *Working Paper*, June 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.

Greenstone, Michael, Richard A. Hornbeck, and Enrico Moretti, "Identifying Agglomeration Spillovers: Evidence from Winners and Losers of Large Plant Openings," *Journal of Political Economy*, 2010, *118* (3), 536–598.

Hombert, Johan and Adrien Matray, "Technology Boom, Labor Reallocation, and Human Capital Depreciation," *Paris December 2018 Finance Meeting EUROFIDAI - AFFI*, February 2019.

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.

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.

Nakamura, Emi and Jón Steinsson, "Fiscal Stimulus in a Monetary Union: Evidence from US Regions," *American Economic Review*, 2014, *104* (3), 753–792.

Perihan, Andrea Weber Saygin and Michele Weynandt, "Coworkers, Networks, and Job Search Outcomes," *ILR Review*, 2014.

Pierce, Justin R. and Peter K. Schott, "Trade Liberalization and Mortality: Evidence from US Counties," *American Economic Review: Insights*, 2018, *forthcoming.*

Powell, Warren, *Approximate Dynamic Programming: Solving the Curses of Dimensionality*, second ed., New York: John Wiley and Sons, 2011.

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.

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 3

A.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.[32]

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.[33]

A.2 Additional Descriptive Statistics on Migration Patterns

[32]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.

[33]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.

B Appendix to Section 2

B.1 Proofs

B.1.1 Proof of Proposition 1

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:

$$\log \frac{\bar{\pi}_k(\pi)(\theta)}{\bar{\pi}_k(\pi)(\hat{\theta})} = g_k \left(\log \frac{\pi_0(\theta)}{\pi_0(\hat{\theta})}, \ldots, \log \frac{\pi_1(\theta)}{\pi_1(\hat{\theta})} \right). \tag{B.1}$$

for all pairs of states $\theta \neq \hat{\theta}$ and all profiles of beliefs π. Furthermore, label neutrality guarantees that the function g_k is independent of θ and $\hat{\theta}$.

Now, consider three distinct states θ, $\hat{\theta}$ and $\tilde{\theta}$. Given that (B.1) 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 (B.1) one more time implies that

$$\log \frac{\bar{\pi}_k(\pi)(\theta)}{\bar{\pi}_k(\pi)(\hat{\theta})} = \int_I a_{ik} \log \frac{\pi_i(\theta)}{\pi_i(\hat{\theta})} di, \quad \forall \theta, \hat{\theta} \in \Theta.$$

Monotonicity implies that $a_{ik} > 0$ for all $i \in I$. Since I assume that the weight that every agents place on each other individual's belief is $a_{ik} = 1$, we can aggregate all beliefs that are identical and write:

$$\log \frac{\bar{\pi}_k(\pi)(\theta)}{\bar{\pi}_k(\pi)(\hat{\theta})} = \sum_j \sum_s L_{jk|s} \log \frac{\pi_{jk|s}(\theta)}{\pi_{jk|s}(\hat{\theta})}. \tag{B.2}$$

B.1.2 Proof of Lemma 1

B.1.3 Proof of Proposition 2

B.1.4 Proof of Proposition 3

$$\pi_{jkt}(\theta_t|\varepsilon_t) = \frac{p_{jk}(\theta_t, L_{t-1}, \varepsilon_t)\pi_{jt}(\theta_t|L_{t-1})}{q_{jk}(L_{t-1}, \varepsilon_t)}. \tag{B.3}$$

$$\log \bar{\pi}_{kt}(\theta_t|L_t) = C_{kt} + \sum_j L_{jt-1} \int_\varepsilon p_{jk}(\theta_t, L_{t-1}, \varepsilon) \log \pi_{jkt}(\theta_t|\varepsilon) d\varepsilon, \tag{B.4}$$

$$\pi_{kt+1}(\theta_{t+1}|L_t) = \int_\theta \bar{\pi}_{kt}(\theta|L_t)\gamma(\theta_{t+1}|\theta)d\theta, \tag{B.5}$$

$$L_{kt} = \sum_j L_{jt-1}\bar{p}_{jk}(\theta_t, L_{t-1}). \tag{B.6}$$

B.1.5 Proof of Proposition 4

B.2 Extension to Internal Trade in Goods

B.3 Extension to Unobserved Population

B.4 Extension to Amenity Shocks

B.5 Extension to Differentiated Networks

B.6 Test of Locally Invariant Posteriors

B.7 Model with Gaussian Correlated Beliefs

B.8 Results for Alternative Values of the Labor Share

C Appendix to Section 4

C.1 Algorithm for Simulated Method of Moments

C.2 Iterative Estimation of the Model with No Information Frictions

C.3 Construction of the Instrument for Internet Access

D Appendix to Section 5

D.1 Fit of Migration Flows

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

D.3 List of Local Shocks

D.4 Replicate Local Shocks in the Model

E Additional Figures

E.1 Cardell Distribution

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

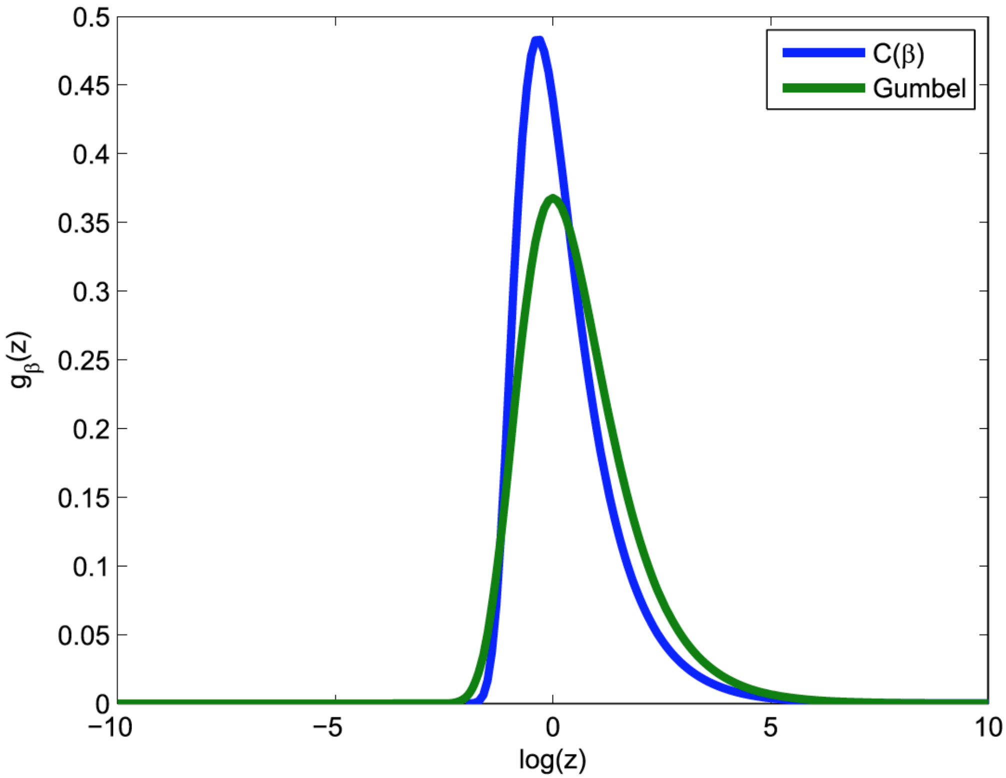


Figure E.1: Distribution Cardell with dispersion $\beta = 0.5$, vs. Gumbel. Source: Dasgupta and Mondria (2018).