ifoCEMIR CESifo

JUNIOR ECONOMIST WORKSHOP ON MIGRATION RESEARCH 2019

Munich, 4–5 July 2019

Rational Inattention and Migration Decisions

Simone Bertoli, Jesús Fernández-Huertas Moraga, and Lucas Guichard

ifoCEMIR
Center of Excellence
for Migration and Integration Research

20 YEARS CESifo

Rational inattention and migration decisions*

Simone Bertoli[a], Jesús Fernández-Huertas Moraga[b], and Lucas Guichard[a,c]

[a] *Université Clermont Auvergne, CNRS, IRD, CERDI*[†]
[b] *Universidad Carlos III de Madrid*[‡]
[c] *IAB*[§]

Abstract

We analyze the implications of introducing a cost related to acquiring information on the attractiveness of the alternative destinations in the decision problem that migrants face. This extension of the canonical model entails that individuals with stronger priors about the identity of their utility-maximizing alternative rationally gather less information. The theoretical model gives us an analytical expression for the expected value of information that can be computed from past migration data. The econometric analysis reveals that migration flows originating from countries characterized by stronger priors are significantly less responsive to variations in economic conditions at destination.

Keywords: international migration; information; gravity equation.

JEL codes: F22; D81; D83.

---

[*] The Authors are grateful to Rabah Arezki, Erhan Artuç, Vianney Dequiedt, Lionel Fontagné, Joël Machado, Thierry Mayer, David McKenzie, Çağlar Özden, Panu Poutvaara, Ariell Reshef, Victor Stephane, Jérôme Valette, and to the participants at seminars at CERDI, the World Bank, Université Paris 1 Panthéon-Sorbonne, IAB and to the IZA-CREA Workshop on "Gravity equations in International Economics: When Trade Meets Migration" (University of Luxembourg) for their comments; the Authors are also grateful to Sergio Correia, Paulo Guimarães and Thomas Zylkin for sharing the Stata command `ppmlhdfe` with us before it became publicly available, and to Olivier Santoni fo providing valuable research assistance; Simone Bertoli acknowledges the support received from the *Agence Nationale de la Recherche* of the French government through the program "*Investissements d'avenir*" (ANR-10-LABX-14-01), and from the *Institut Universitaire de France*; Jesús Fernández-Huertas Moraga acknowledges the financial support from the *Ministerio de Economía, Industria y Competitividad* (Spain), grants ECO2016-76402-R and MDM 2014-0431, and from the *Comunidad de Madrid*, MadEco-CM (S2015/HUM-3444); the usual disclaimers apply.

[†] CERDI, Avenue Léon-Blum, 26, F-63000, Clermont-Ferrand; email: simone.bertoli@uca.fr.

[‡] Madrid, 126, E-28903, Getafe (Madrid); email: jesferna@eco.uc3m.es.

[§] Regensburger Strasse, 104, D-90478 Nuremberg; email: lucas.guichard2@iab.de.

"Before making a choice, one may have an opportunity to study the actions and their payoffs; however, in most cases it is too costly to investigate to the point where the payoffs are known with certainty. As a result, some uncertainty about the payoffs remains when one chooses among the actions even if complete information was available in principle."

(Matějka and McKay, 2015, p. 272)

1 Introduction

The canonical micro-foundations of a gravity equation for international migration draw on discrete choice models *à la* McFadden, which were initially conceived to describe consumption choices, that assume that the decision maker costlessly observes the realizations of the payoff associated to each alternative in the choice set before selecting her preferred option. This analytical choice is not immaterial, as "[c]onsumers are continually making choices among products, the consequences of which they are dimly aware" (Nelson, 1970, p. 311), and this just partial awareness, *a fortiori*, applies to individuals that have to decide where to migrate. The uncertainty surrounding the utility associated to the various countries of destination might not be entirely resolved when a migrant has to come up with a solution to the location-decision problem that she faces,[1] and the size of the remaining uncertainty could be endogenously determined by migrants' choices to refine their knowledge about the actual attractiveness of the various destinations, as the initial quote by Matějka and McKay (2015) suggests.[2] The literature on rational inattention (Sims, 1998, 2003), which has been recently applied to discrete choice situations (Matějka and McKay, 2015; Caplin et al., 2019), provides us with a framework to think about how costs associated to information acquisition and processing would influence the specification of the migration gravity equation that is brought to the data.

Can we sharpen our understanding of the determinants of international migration flows if we take into account the uncertainty that migrants face, and the actions that they can take to narrow it down? We propose an extension of the standard additive random utility maximization model used in the migration literature featuring a rationally inattentive behavior

---

[1] See McKenzie et al. (2013) for empirical evidence on the inaccuracy of the expectations held by potential migrants on their earnings in a foreign labor market.

[2] In his seminal contribution, Simon (1959) observed that information "should be gathered up to the point where the incremental cost of additional information is equal to the incremental profit that can be earned from having it." (pp. 269–70).

of the migrants. The main prediction of this model is that the responsiveness of bilateral migration flows with respect to variations in the attractiveness of alternative destinations increases when migrants have stronger incentives to acquire information before deciding where to move. These incentives are inversely related to the strength of the migrants' prior beliefs about the identity of their preferred destination, and we propose a way to measure this strength using data on the distribution of past bilateral migration flows. The estimation of a gravity equation reveals that variations in economic conditions in a given destination influence more incoming migration flows from origins where migrants (rationally) invest more in information acquisition.

The contribution of our paper is threefold. First, we extend the canonical microfoundations of the gravity equation in the international migration literature to allow for a costly acquisition of information about the attractiveness of the various destination countries, drawing on the recent contributions by Matějka and McKay (2015) and Caplin et al. (2019). We derive an analytical expression for the value of observing the actual individual-specific utility in each alternative in the choice set, which is inversely related to the strength of the priors held by the migrants. Second, we propose an empirical counterpart of the theoretical expression for the value of information, which is given by (minus) the logarithm of the origin-specific share of migrants in the main destination over a period ranging from 5 to 20 years before the one over which we measure migration flows. We show that this variable significantly mediates the responsiveness of observed bilateral migration flows with respect to variations in economic conditions at destination, uncovering a relevant dimension of heterogeneity that had remained so far unnoticed. Third, we provide evidence of the empirical relevance of rational inattention in discrete choice situations, complementing a strand of literature that it is still mostly theoretical.[3] Migrants appear to be rationally inattentive even though the stakes related to their location decisions are certainly very high (see, for instance, McKenzie et al., 2010 and Clemens et al., forthcoming).

We draw on data on bilateral migration flows between 1960 and 2015 from Abel (2018) to build an origin-specific time-varying measure of the value of information for international migrants, whose functional form comes directly from our theoretical model. We estimate a gravity equation where the deterministic component of destination-specific utility is augmented with an interaction between income per capita and the empirical counterpart of our

---

[3] "The model of [rational inattention] is well suited for a boom in empirical work, which has not yet occurred." (Maćkowiak et al., 2018, p. 27).

measure of the value of information. The results are in line with the theoretical model: a one standard deviation increase in our proxy for the value of information determines an increase in the estimated elasticity between 0.063 and 0.083. Our estimates entail that the elasticity of the bilateral migration rate with respect to income per capita for China is 0.160-0.216 higher than the corresponding elasticity for Mexico, which represents a paradigmatic case of migration flows concentrated in just one single destination, the United States. The econometric evidence that we provide is fully robust when we allow for additional heterogeneity of this elasticity at the origin or at the dyadic-level.

Our paper is related to and contributes to three different strands of literature: (*i*) rational inattention in discrete choice models, (*ii*) micro-foundations of the migration gravity equations, and (*iii*) migration models with residual uncertainty.

As far as rational inattention in discrete choice models is concerned, the closest reference to our theoretical model is Matějka and McKay (2015). They consider a discrete choice model where the decision maker chooses the precision of the signals that she receives about the payoffs of the various alternative in the choice set, with the cost of observing the signal being proportional to the ensuing reduction in the Shannon or differential entropy associated to the distribution of payoffs (Shannon, 1948). Matějka and McKay (2015) demonstrate that choice probabilities depend on the actual (but possibly *un*observed) values of the payoffs, and on the prior beliefs that individuals have about the attractiveness of each alternative in the choice set, which shape the chosen information-processing strategy. The decision maker could rationally decide to completely disregard some of the alternatives in the choice set (Caplin et al., 2019), thus choosing with a positive probability in any state of the world only a subset of options belonging to the so-called consideration set, something that would never occur if information was costlessly available. Our reliance on the distribution of past migration flows across destinations to measure the value of information acquisition is closely related to the use of past market shares in Caplin et al. (2016). Dasgupta and Mondria (2018) have drawn on Matějka and McKay (2015) to extend the Ricardian model of trade by Eaton and Kortum (2002), allowing for a costly acquisition of information on the prices of a good in different countries. The main prediction of their model refers to the non-monotonicity of the relationship between the number of distinct countries from which the same good is imported and the unitary cost of information, with the data being consistent with the predicted hump-shaped relationship between the two variables.

With respect to micro-foundations of the migration gravity equations, our paper is related to papers that have derived a specification of the gravity equation under more general distributional assumptions on the stochastic component of utility, notably Bertoli and Fernández-Huertas Moraga (2013) and Ortega and Peri (2013), as these more general specifications have allowed uncovering additional determinants of bilateral migration flows. Batista and McKenzie (2018) have tested in the lab these micro-foundations, notably allowing players to pay a cost to reduce the uncertainty about the payoffs associated to the various destinations. Fosgerau et al. (2018) have recently drawn a bridge between the first two strands of literature that our paper is related to. Specifically, they demonstrate that choice probabilities in Matějka and McKay (2015) follow a modified logit formula because of the reliance of Shannon entropy to measure the cost of information acquisition.[4] Relying on a generalized entropy function, which allows for spillovers in information acquisition across alternatives, results in choice probabilities that can be derived from *any* additive random utility maximization model.

Finally, our paper is related to models of migration with unresolved uncertainty that arises either because utility at destination is not remotely observable (Bertoli, 2010), or because of the repeated nature of the location-decision problem that migrants face (Kennan and Walker, 2011; Bertoli et al., 2016; Artuç and Özden, 2018), with location choices that can be revised as individuals draw new realizations of the stochastic components of utility in each period.[5]

The rest of the paper is structured as follows: Section 2 combines the canonical additive random utility maximization model with rational inattention arising from a costly acquisition of information about the attractiveness of the various destination countries; Section 3 briefly presents the main data sources, it describes how we bring the testable implication of the theoretical model to the data, and it presents basic descriptive statistics. Section 4 presents the results of the estimation of the migration gravity equation that allows for rationally inattentive location choices, and Section 5 concludes.

---

[4] See also Caplin et al. (2017) for a characterization and a generalization of Shannon entropy in models of rational inattention.

[5] Steiner et al. (2017) include rational inattention in a dynamic discrete choice model.

2 Theoretical model

Let u_{ijk} denote the utility that migrant i from the origin j derives if she opts for country $k \in A$, where A represents her choice set, with $\#A = N$, which does not include country j itself, so that we focus here on the choice of the destination conditional upon migrating. Assume that u_{ijk} is additive in a deterministic component of utility, v_{ijk}, and in a stochastic component, ϵ_{ijk}:[6]

$$u_{ijk} \equiv v_{ijk} + \epsilon_{ijk} \tag{1}$$

Let $\boldsymbol{v}_{ij} = (v_{ij1}, v_{ij2}, ..., v_{ijN})'$, $\boldsymbol{\epsilon}_{ij} = (\epsilon_{ij1}, \epsilon_{ij2}, ..., \epsilon_{ijN})'$, and $\boldsymbol{e}_{ij} = (e_{ij1}, e_{ij2}, ..., e_{ijN})'$ represent three $N \times 1$ column vectors stacking respectively the deterministic component of utility, the stochastic component of utility and its realizations across all alternatives in the choice set A. The standard micro-foundations of this location-decision problem are based on distributional assumptions *à la* McFadden (1974), with:

$$F_{\boldsymbol{\epsilon}}(\boldsymbol{\epsilon}_{ij}) = e^{-\sum_{k \in A} t(\epsilon_{ijk})} \tag{2}$$

where $t(\epsilon_{ijk}) = e^{-(\epsilon_{ijk}+\gamma)}$ and $\gamma \approx 0.5772$ is Euler's constant,[7] as an identically and independently distributed EVT-1 stochastic component of utility allows to derive a specification a for gravity equation that can be easily estimated with aggregate data.[8]

The stochastic component of location-specific utility in (1) can reflect both individual heterogeneity in preferences and aggregate uncertainty about the attractiveness of alternative $k \in A$, while maintaining the distributional assumptions described in (2). Specifically, we could define:

$$\epsilon_{ijk} \equiv C_{jk}(\alpha_k) + \alpha_k \eta_{ijk}, \tag{3}$$

with $\alpha_k \in (0, 1]$, η_{ijk} following an EVT-1 distribution, and $C_{jk}(\alpha_k)$ being the (unique) random variable that ensures that ϵ_{ijk} also follows an EVT-1 distribution (Cardell, 1997). If $\alpha_k = 1$, then C_{jk} is equal to an arbitrary constant, and it is not stochastic, so that (3) also

---

[6] The cost of moving from the origin j to the destination k is included into the dyadic deterministic component of utility.

[7] The inclusion of γ in the expression for $t(\epsilon_{ijk})$ ensures that the expected value of the stochastic component of utility ϵ_{ijk} is equal to 0.

[8] This occurs because the dependence of bilateral migration flows on the attractiveness of alternative destinations in the choice set can be controlled through a suitable normalization of the dependent variable, as it occurs in the Ricardian model of trade by Eaton and Kortum (2002), while the same would not occur with a normally distributed stochastic component, as in Roy (1951) or Borjas (1987).

nests the case in which the stochastic component of utility only reflects unobserved individual heterogeneity. $C_{jk}(\alpha_k)$, which does not vary across individuals, captures the aggregate uncertainty for migrants from j about the attractiveness of alternative $k \in A$, and the relative importance of individual heterogeneity and aggregate uncertainty, described by α_k, could be destination-specific. Assuming that $C_{jk}(\alpha_k)$ and η_{ijk} are independently distributed across alternatives in the choice set ensures that ϵ_{ijk} is consistent with the distributional assumptions laid out in (2). The inclusion of $C_{jk}(\alpha_k)$ in (3) entails that, as in Matějka and McKay (2015), the migrant faces an uncertainty about the state of nature that influences location-specific utility, while the individual-specific stochastic component η_{ijk} might be perfectly known to her. This entails that our framework is fully consistent with a classical interpretation of the stochastic component of utility, which reflects the imperfect knowledge of the modeler.[9] Thus, we do not need to assume here, as in Caplin et al. (2016), that the decision maker is uncertain about her own preferences over the various alternatives in the choice set.

2.1 Migrants' information set

We analyze this location-decision problem under two alternative assumptions on the information set to which migrants have access before choosing their preferred destination: the canonical full information case, where migrants observe both $\boldsymbol{v}_{ij}$ and $\boldsymbol{e}_{ij}$,[10] and a partial information case, where migrants only observe $\boldsymbol{v}_{ij}$. This second more limited information set is meant to represent a (limit) case of the constraints on the capacity to process information that could characterize migrants' decisions, and it will help us to understand what shapes the incentives to acquire and process information.

2.1.1 Full information

Let us define the real-valued function $V(\boldsymbol{v}_{ij} + \boldsymbol{e}_{ij})$ that returns the utility associated to the utility-maximizing alternative in A. Formally, $V : \mathbb{R}^N \rightarrow \mathbb{R}$, with $V(\boldsymbol{v}_{ij} + \boldsymbol{e}_{ij}) \equiv$

---

[9] "The econometricians' approach is conceptually very different [...] both the decision rule and the utility functions of the individual are deterministic. *The uncertainty is due to the lack of information available to the modeler.*" (Anderson et al., 1992, pp. 31–33, emphasis in the text).

[10] Borjas (1987) assumes that migration decisions are based on a comparison of "potential incomes" at origin and at destination (p. 532), with the latter being remotely observable, i.e., known before migrating, in line with the analysis by Roy (1951) on the occupational choice between hunting and fishing that explicitly assumes that "[e]very man, too, has a fairly good idea of what his annual output is likely to be in both occupations" (p. 137).

$\max_{k \in A}(v_{ijk} + e_{ijk})$. Similarly, we can define $a : \mathbb{R}^N \rightarrow A$ as the alternative in the choice set A such that $(v_{ija} + e_{ija}) = \max_{k \in A}(v_{ijk} + e_{ijk})$. Under the distributional assumptions that we introduced above, the *ex ante* probability that alternative k is the utility-maximizing alternative for individual i is given by (McFadden, 1974, 1978):[11]

$$p_{ijk} = \frac{e^{v_{ijk}}}{\sum_{l \in A} e^{v_{ijl}}} \tag{4}$$

Notice that the decision rule of individual i is not probabilistic, as the vector $\boldsymbol{e}_{ij}$ is observed before deciding where to migrate; by the law of large numbers, p_{ijk} coincides with the share of i-type individuals from the origin j that find it optimal to migrate to k. The expected utility from the choice situation is given by the integral, over the distribution of $\boldsymbol{\epsilon}_{ij}$, of $V(\boldsymbol{v}_{ij} + \boldsymbol{\epsilon}_{ij})$, the utility associated to the utility-maximizing alternative. Under the distributional assumptions introduced above, we have that (Ben-Akiwa and Lerman, 1979; Small and Rosen, 1981):

$$\mathbb{E}_{\boldsymbol{\epsilon}_{ij}}[V(\boldsymbol{v}_{ij} + \boldsymbol{\epsilon}_{ij})] = \ln\left(\sum_{l \in A} e^{v_{ijl}}\right) \tag{5}$$

2.1.2 Limited information

What if the decision to migrate has to be taken observing just $\boldsymbol{v}_{ij}$? In this case, migrant i should opt for the alternative characterized by the highest deterministic component of utility. We can define the real-valued function $W(\boldsymbol{v}_{ij})$ that returns the highest deterministic component of utility in the choice set A. Formally, $W : \mathbb{R}^N \rightarrow \mathbb{R}$, with $W(\boldsymbol{v}_{ij}) \equiv \max_{k \in A} v_{ijk}$. Similarly, we can define $c : \mathbb{R}^N \rightarrow A$ as the alternative in the choice set A such that $v_{ijc} = \max_{k \in A} v_{ijk}$. The expected utility from opting for alternative $c \in A$ is simply given by:

$$\mathbb{E}_{\boldsymbol{\epsilon}_{ij}}[W(\boldsymbol{v}_{ij})] = \int_{-\infty}^{+\infty} u_{ijc}\, \mathrm{d}\epsilon_{ijc} = v_{ijc} \tag{6}$$

2.2 Rational inattention

Let us consider the case in which the realizations of the stochastic component of utility can be observed by individual i, but this observation comes at a cost. In line with the literature on rational inattention (Matějka and McKay, 2015), we assume that the cost of acquiring

---

[11] Luce and Suppes (1965, p. 338) credit unpublished work by Holman and Marley for this result.

information on the actual values of location-specific utility is proportional to the entropy of the multivariate distribution of utility, i.e., the vector $\boldsymbol{v}_{ij} + \boldsymbol{\epsilon}_{ij}$. We derive next the cost of a (discrete) switch from the partial to the full information case described above, and we also derive the expected value of acquiring (full) information on the values of the stochastic components of utility.

2.2.1 The cost of acquiring information

Let B denote the belief of migrant i about the distribution of the stochastic component of utility, and let us assume that B coincides with the actual distribution described in (2). The differential entropy $H(B)$ associated to this distribution is given by:

$$H(B) \equiv -\int_{\boldsymbol{\epsilon}_{ij}} f(\boldsymbol{\epsilon}_{ij}) \ln f(\boldsymbol{\epsilon}_{ij}) \, \mathrm{d}\boldsymbol{\epsilon}_{ij} \tag{7}$$

where $f(\boldsymbol{\epsilon}_{ij})$ is the probability density function corresponding to the cumulative density function in (2). The joint entropy of N independent distributions is given by the sum of their individual entropies (see, for instance, Cover and Thomas, 1991), while the assumption that the distributions are identical entails that the joint entropy in (8) is given by $H(B) = \sum_{k \in A} H(\epsilon_{ijk})$; the entropy $H(\epsilon_{ijk})$ of a univariate distribution is equal to:

$$H(\epsilon_{ijk}) = -\int_{-\infty}^{+\infty} f(\epsilon_{ijk}) \ln f(\epsilon_{ijk}) \, \mathrm{d}\epsilon_{ijk} \tag{8}$$

With some simple algebraic manipulations,[12] we have that:

$$\begin{aligned} H(\epsilon_{ijk}) &= \int_{-\infty}^{+\infty} [t(\epsilon_{ijk}) + (\epsilon_{ijk} + \gamma)] f(\epsilon_{ijk}) \, \mathrm{d}\epsilon_{ijk} \\ &= \int_{-\infty}^{+\infty} t(\epsilon_{ijk}) f(\epsilon_{ijk}) \, \mathrm{d}\epsilon_{ijk} + \gamma \\ &= [t(\epsilon_{ijk}) F(\epsilon_{ijk})]_{-\infty}^{+\infty} + \int_{-\infty}^{+\infty} t(\epsilon_{ijk}) F(\epsilon_{ijk}) \, \mathrm{d}\epsilon_{ijk} + \gamma \\ &= [f(\epsilon_{ijk})]_{-\infty}^{+\infty} + \int_{-\infty}^{+\infty} f(\epsilon_{ijk}) \, \mathrm{d}\epsilon_{ijk} + \gamma = 1 + \gamma \end{aligned} \tag{9}$$

---

[12] We use the fact that $f(\epsilon_{ijk}) = t(\epsilon_{ijk}) F(\epsilon_{ijk})$, and we rely on integration by parts to integrate $t(\epsilon_{ijk}) f(\epsilon_{ijk})$, with the indefinite integral of $t(\epsilon_{ijk})$ being $-t(\epsilon_{ijk})$.

Letting $\lambda_k \in \mathbb{R}^+$, with $k \in A$, denote the destination-specific parameter that converts entropy into the metric of utility, the cost $c(\lambda_A)$ of switching from partial to full information is given by:

$$c(\lambda_A) = \lambda_A(1+\gamma) \tag{10}$$

where $\lambda_A \equiv \sum_{k \in A} \lambda_k$.

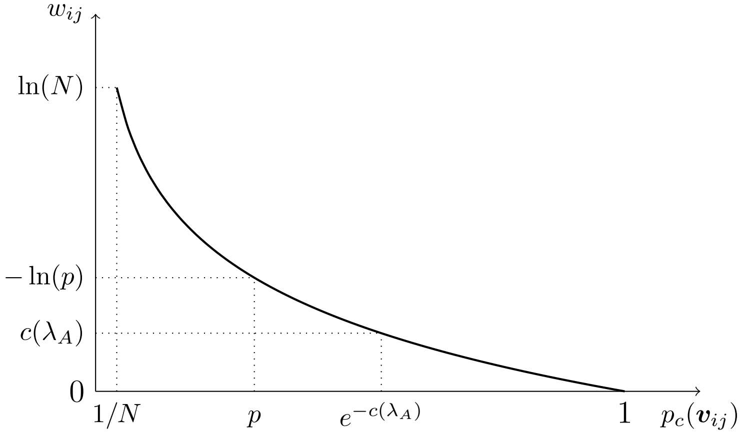
2.2.2 The value of acquiring full information

The value w_{ij} of observing the realization of the stochastic component of utility is given by the difference between the value of the choice situation under full and partial information, i.e., $\mathbb{E}_{\boldsymbol{\epsilon}_{ij}}[V(\boldsymbol{v}_{ij} + \boldsymbol{e}_{ij})] - \mathbb{E}_{\boldsymbol{\epsilon}_{ij}}[W(\boldsymbol{v}_{ij})]$. From (5) and (6), we have that w_{ij} is equal to:

$$w_{ij} \equiv \mathbb{E}_{\boldsymbol{\epsilon}_{ij}}[V(\boldsymbol{v}_{ij} + \boldsymbol{e}_{ij})] - \mathbb{E}_{\boldsymbol{\epsilon}_{ij}}[W(\boldsymbol{v}_{ij})] = \ln\left(\sum_{k \in A} e^{v_{ijk}}\right) - v_{ijc} = -\ln[p_c(\boldsymbol{v}_{ij})] \tag{11}$$

where $p_c(\boldsymbol{v}_{ij}) \in [1/N, 1]$ is the probability that the destination with the highest deterministic component of utility has also the highest utility, i.e., $p_c(\boldsymbol{v}_{ij}) = \text{Prob}\,[c(\boldsymbol{v}_{ij}) = a(\boldsymbol{v}_{ij} + \boldsymbol{e}_{ij})]$.

Figure 1: The value of full information w_{ij}



The value of full information is a monotonically decreasing function of this probability, as shown in Figure 1. Intuitively, the acquisition of information increases the expected utility derived from the choice situation only if it leads to a change in the alternative in the choice set that is selected by the migrant. The higher is $p_c(\boldsymbol{v}_{ij})$, the lower are the chances that a (costly) acquisition of information would induce the migrant to modify the location choice based on her priors. We can refer to $p_c(\boldsymbol{v}_{ij})$ as the strength of migrant's priors, based on exclusively $\boldsymbol{v}_{ij}$, concerning the identity of the utility-maximizing destination in the choice set.

2.3 Strength of the priors and return from acquiring information

The value of acquiring information is negatively related to the strength of the priors held by migrant i, while the cost in terms of utility of such an investment is *in*dependent from the strength of the priors. Thus, the expected return $\pi_{ij}(\boldsymbol{v}_{ij}, \lambda_A)$ from acquiring information, given by the difference between (11) and (10), is equal to:

$$\pi_{ij}(\boldsymbol{v}_{ij}, \lambda_A) = w_{ij} - c(\lambda_A) = -\ln[p_c(\boldsymbol{v}_{ij})] - \lambda_A(1+\gamma) \tag{12}$$

When information acquisition is costly, migrants with stronger priors will tend to be rationally inattentive, as the acquisition of information about the actual attractiveness of the various destinations in the choice set is unlikely to induce them to modify their destination choices, and it is thus associated with a lower value of $\pi_{ij}(\boldsymbol{v}_{ij}, \lambda_A)$. Furthermore, (12) implicitly defines a threshold value of $p_c(\boldsymbol{v}_{ij})$, depicted in Figure 1, such that migrants will certainly not find profitable to acquire full information when $p_c(\boldsymbol{v}_{ij}) > p(\lambda_A) \equiv e^{-c(\lambda_A)}$ even if this was potentially available, as the initial quote from Matějka and McKay (2015) suggests.[13]

Migrants are not constrained to select one of the two polar cases that we have analyzed, as they could opt for observing Bayes-consistent signals about location-specific utility (Matějka and McKay, 2015), possibly gathering information just on a subset of the choice set A, rationally disregarding alternatives that are, *a priori*, extremely unlikely to be the utility-maximizing option (Caplin et al., 2019). Rational inattention results in choice probabilities that have a structure that resembles a canonical conditional logit but that depend both on the actual (but possibly unobserved) attractiveness of each alternative, and on the prior beliefs held by the decision maker.[14]

---

[13] Notice that $p_c(\boldsymbol{v}_{ij}) > p(\lambda_A)$ represents a sufficient but not a necessary condition for migrants rationally deciding not to acquire full information on the attractiveness of each alternative in the choice set A.

[14] Appendix A.1 reports the choice probabilities that solve the location-decision problem with costly information acquisition.

This gives us a clear testable implication: if the stochastic component of location-specific utility also includes an aggregate component, as in (3), and the actual investment in information acquisition is proportional to the return from acquiring full information in (12),[15] then migration flows should be less responsive to variations in the attractiveness of the destination countries when migrants have stronger priors. Rationally inattentive migrants with strong priors (optimally) decide to gather less information about the attractiveness of the various destinations, and hence their location choices are less sensitive to variations in, say, economic conditions than we would expect when full information can be costlessly acquired. The theoretical model does not just give us a qualitative prediction, but also very specific guidelines with respect to the functional form, described in (12), of the relationship between the return from acquiring information and the strength of migrants' priors.

3 From the theory to the data

We describe here the source of our panel data on bilateral international migration flows, and how we build from these data the empirical counterpart of the theoretical definition of the value of information. We also present basic descriptive statistics, focusing in particular on our variable of interest.

3.1 Data on bilateral migration flows

Our main data source is represented by Abel (2018), which provides data on the bilateral migration flows $m_{jkt} \geq 0$ between the origin j and the destination k across 203 countries for five-year periods, starting in t, between 1960 and 2015. Abel (2018) extends the methodology presented by Abel and Sander (2014) for inferring gender-specific bilateral migration flows from information on the stock of individuals (by country of birth) residing in each country obtained from population censuses. More precisely, Abel (2018) recovers the minimal amount of bilateral flows that are required in order to match the observed evolution of stock data, which are adjusted for demographic events. The stock data are taken from Özden et al. (2011) between 1960 and 2000, and from United Nations Population Division (2015a) for later years,

---

[15] The Appendix A.2 shows that, in a simplified version of the model that is analytically tractable, the actual investment in information acquisition is bounded from above by a term that is proportional to the return from acquiring full information.

and are combined with demographic information from United Nations Population Division (2015b) to obtain the estimates on flows. To our knowledge, the dataset generated by Abel (2018) is the most comprehensive in terms of both time and geographical coverage produced to date on international migration flows.[16] As discussed below, these two aspects are critical to generate from the data the empirical counterpart of the value of information described in our theoretical model. The sample over which we conduct our econometric analysis includes the entire set of countries covered by Abel (2018): for the period between 1980 and 2015, we have 263,008 observations on bilateral migration flows over seven consecutive five-year periods.[17,18] The average value of m_{jkt} stands at 957.4, with a standard deviation of 15,472.4 and a share of zero flows equal to 61.2 percent.

3.2 Measurement of the value of information

The theoretical model predicts that the return from information acquisition in (12) depends on the strength of migrants' priors concerning the (time-varying) identity of the utility-maximizing destination, and on the cost of acquiring information. This requires building from the data a suitable empirical counterpart of the value of information, and dealing in the econometric analysis with the challenges related to the *un*observability of the cost of information. Our assumption is that current migrants from j have observed the destination choices made in the past by individuals leaving from j itself, and their distribution across destinations has shaped the strength of their priors. The location-decision problem presented in Section 2 is static, and it allows for individual heterogeneity; the availability of longitudinal data on bilateral migration flows that are not disaggregated (except for the gender dimension) entails here that we have to assume that choice probabilities in (4) are constant across individuals for a given origin but possibly time-varying. Thus, we proxy the time-varying strength of the priors for the origin country j in year t with the share of migration flows directed from j to the main foreign destination in a period up to t. More precisely, we rely

---

[16] Our empirical evidence is robust to using only the bilateral flow data in Abel (2018) that are based solely on migrant stocks from Özden et al. (2011), thus avoiding possible inconsistencies at the junction between the two underlying data sources, and to defining bilateral migration flows as the variations in the stock of j-born individuals residing in destination k derived from Özden et al. (2011).

[17] Migration flows before 1980 are used to measure the empirical counterpart of the value of information, as discussed in Section 3.2 below.

[18] This is below $203 \times 202 \times 7 = 287,042$ as we have missing information of GDP per capita at destination for some destination-year pairs; more precisely, we lose completely 14 destination countries, which represent less than 0.9 percent of total migration flows in Abel (2018).

on $p(r)_{jt}$, defined as follows:[19]

$$p(r)_{jt} \equiv \max_k \left\{ \frac{\sum_{t-r}^{t} m_{jkt}}{\sum_{t-r}^{t} \sum_{l \in A} m_{jlt}} \right\}, \; r = \{5, 10, 15, 20\} \tag{13}$$

It is interesting to note that 107 different countries represent the main destination, and hence determine the value of information, for at least one of the 1,349 origin-year pairs in our estimation sample; intuitively, the United States are the most typical main destination accumulating most of the flows for a particular origin, but this happens only in 20.7 percent of the cases; the second most typical main destination is Russia, for 7.6 percent of all origin-year pairs, and five Sub-Saharan African countries (namely, South Africa, Ethiopia, Nigeria, the Democratic Republic of the Congo, and Ivory Coast) appear among the 20 countries that most frequently play the role of main destination. We then measure, following (11), the value of information for the origin j at time t as:

$$w(r)_{jt} = -\ln[p(r)_{jt}] \tag{14}$$

To give concrete examples, we have that 97.0 percent of flows from Mexico between 1990 and 1995 were directed to the United States, so that $w(5)_{\text{MEX1995}} = -\ln(0.970) = 0.031$. Over the same period, 25.4 percent of migration flows from China were directed to the main destination (United States), and this entails that $w(5)_{\text{CHN1995}} = -\ln(0.254) = 1.371$. Thus, the empirical counterpart of the value of information in (14) entails that Chinese migrants valued information more than Mexican migrants in the five-year period starting in 1995, as the latter group of migrants held substantially stronger priors with respect to the identity of the utility-maximizing destination.

---

[19] Notice that $p(r)_{jt}$ in (13) is defined provided that the total flow originating from j between year $t-r$ and t is positive; this is always the case except for 31 origin-year pairs when $r = 5$, 14 origin-year pairs when $r = 10$, 7 when $r = 15$, and 6 when $r = 20$.

Table 1: Descriptive statistics for the empirical counterparts of the value of information

	mean	s.d.	min	max	obs.
$w(5)_{jt}$	0.86	0.53	0.00	2.49	257,086
$w(10)_{jt}$	0.92	0.52	0.00	2.40	260,332
$w(15)_{jt}$	0.95	0.52	0.00	2.53	261,668
$w(20)_{jt}$	0.96	0.52	0.00	2.47	261,858

Notes: $w(r)_{jt}$, with $r = \{5, 10, 15, 20\}$, computed according to (14).

Source: Authors' elaboration on Abel (2018).

Going beyond specific examples, Table 1 reports the descriptive statistics for $w(r)_{jt}$, with $r = \{5, 10, 15, 20\}$. The average value of the empirical counterpart of the value of information monotonically increases with r, from 0.86 for $w(5)_{jt}$ to 0.96 for $w(20)_{jt}$, as the share of migrants from j directed to the main destination declines with the length of the period over which we measure past migration flows. Nevertheless, the four variants of the empirical counterparts of the value of information are closely correlated. Table 2 shows the exact numbers for our baseline dataset: as expected, all the measures are positively correlated but the correlations range between 0.58 between $w(5)_{jt}$ and $w(20)_{jt}$, and 0.93 for $w(15)_{jt}$ with $w(20)_{jt}$.

Table 2: Correlation among the proxies for the value of information

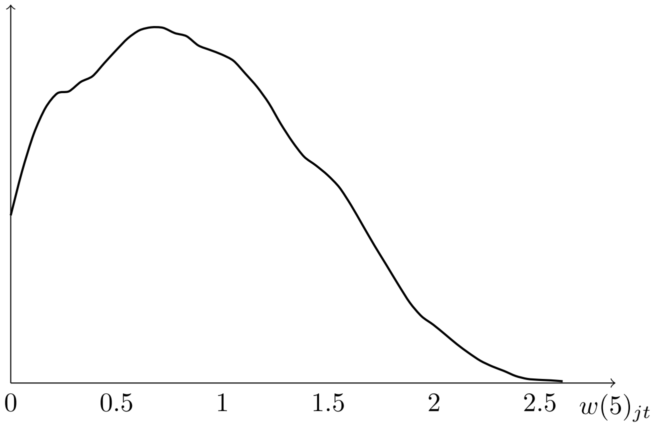
	$w(5)_{jt}$	$w(10)_{jt}$	$w(15)_{jt}$	$w(20)_{jt}$
$w(5)_{jt}$	1.00			
$w(10)_{jt}$	0.77	1.00		
$w(15)_{jt}$	0.65	0.86	1.00	
$w(20)_{jt}$	0.58	0.76	0.93	1.00

Notes: $w(r)_{jt}$, with $r = \{5, 10, 15, 20\}$, computed according to (14).

Source: Source: Authors' elaboration on Abel (2018).

Figure 2 plots the smoothed distribution of $w(5)_{jt}$; observed values for $w(5)_{jt}$ range between 0 and 2.49, as reported in Table 1, thus covering a substantial portion of the range of values that are theoretically feasible.[20] The variability in Figure 2 reflects both time-invariant differences across origins, as well as within-origin differences over time. More precisely, a regression of $w(5)_{jt}$ on a set of origin dummies explains only 40.4 percent of the variability in the value of information.

Figure 2: Kernel distribution for the value of information $w(5)_{jt}$



Source: Authors' elaboration on Abel (2018).

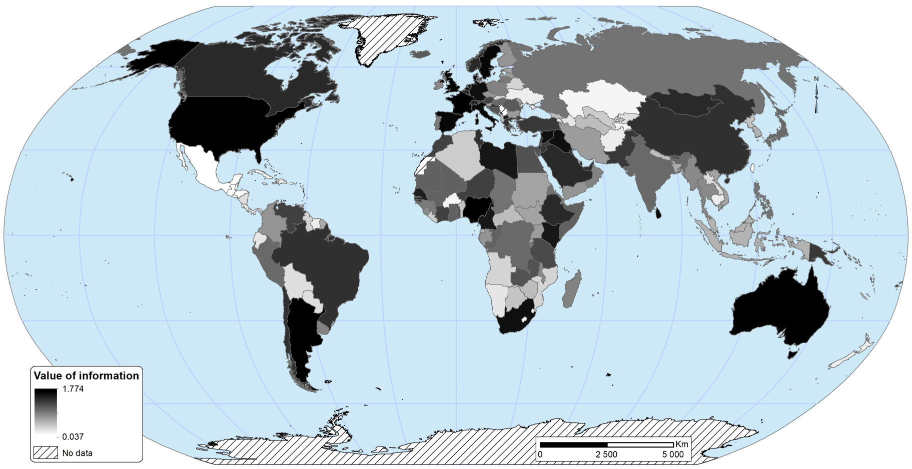
Figure 3 plots the origin-specific average of the value of information $w(5)_{jt}$ between 1980 and 2015 on a world map, revealing that there is no clear geographical pattern in the data, with a substantial variability in the value of $w(5)_{jt}$ within, say, Latin America or Sub-Saharan Africa.[21] Figure 3 also reveals that high-income countries in Western Europe, North America and Oceania are typically characterized by a high average value of $w(5)_{jt}$, a pattern that will be taken into account in the econometric analysis.

---

[20] With $N = 184$, the upper bound of the value of information stands at $\ln(184) \approx 5.2$.

[21] Notice that the range of values for $w(5)_{jt}$ is narrower in Figure 3 than in Table 1 as in the former we are averaging across periods for each origin.

Figure 3: Origin-specific average of the value of information $w(5)_{jt}$



Source: Authors' elaboration on Abel (2018).

The theory entails that the empirical counterpart $\pi(r)_{jt}$, with $r = \{5, 10, 15, 20\}$, of the return from information acquisition in (12) is simply given by the difference between the empirical counterpart $w(r)_{jt}$ of the value of information and an (unknown) constant term. This, in turn, implies that the evidence that we have just provided on the distribution of $w(5)_{jt}$, on its correlation with the proxies obtained using past migration flows over longer time periods, and on the relevance of within-origin variability applies also to the proxy for the return from information acquisition.

4 Econometric analysis

In this section, our objective is to test the empirical relevance of the value of information in shaping migration decisions. To this end, we will augment a canonical gravity specification of the determinants of migration flows with interactions of the empirical value of information and income per capita at destination. The hypothesis that we will test, following our theoretical model, is whether migrants respond more to changes in economic conditions at destination when migrants attach a higher value to information. Hence, we expect the interaction between $w(r)_{jt}$ and income per capita in the destination country to be positive and significant.

4.1 Gravity equation with rational inattention

As a starting point, let us assume that information on the individual-specific values of location-specific utilities can be costlessly acquired, i.e., $\lambda_A = 0$. In this case, we could write the migration flows m_{jkt} between an origin j and a destination k in the five-year period starting in year t as:

$$m_{jkt} = p_{jkt} \times n_{jt} \times \zeta_{jkt} \tag{15}$$

where n_{jt} is the population residing in country of origin j in year t, ζ_{jkt} is an error term, and the probability that destination k represents the utility-maximizing alternative for a migrant from j is given by:

$$p_{jkt} = \frac{e^{v_{jkt}}}{\sum_{l \in A} e^{v_{jlt}}} \tag{16}$$

Replacing p_{jkt} with the expression in (16), we can then rewrite equation (15) as:

$$m_{jkt} = \exp\left[v_{jkt} - \ln\left(\sum_{l \in A} e^{v_{jlt}}\right) + \ln(n_{jt}) + \ln(\zeta_{jkt})\right] \tag{17}$$

We assume that the deterministic component of utility v_{jkt} in (17) follows:

$$v_{jkt} = \alpha \ln\left(\frac{y_{kt}}{\tau_{jkt}}\right) \tag{18}$$

where y_{kt} is real GDP per capita in destination k in year t, and $\tau_{jkt} \geq 1$ are dyadic and time-varying iceberg migration costs. The specification in (18) entails that the semi-elasticity of v_{jkt} with respect to y_{kt} is always equal to α, and independent of the value of the determinants of dyadic migration costs τ_{jkt}. We specialize these dyadic costs as follows:

$$\ln(\tau_{jkt}) = d_{kt} + d_{jt} + d_{jk} + \theta \ln(s_{jkt} + 1) \tag{19}$$

where d_{kt}, d_{jt} and d_{jk} represent destination-time, origin-time and origin-destination (dyadic) dummies, and where $\ln(s_{jkt} + 1)$ represents the logarithm of (one plus) the stock of j-born migrants residing in destination k in year t, as in Beine et al. (2011).[22] The rich structure of fixed effects included in (19) entails that we can only identify the effect of time-varying dyadic variables, such as $\ln(s_{jkt} + 1)$, or of destination-specific time-varying variables that produce heterogeneous effects across origins or origin-destination pairs, while we cannot obtain an estimate for α, which represents the elasticity of the bilateral migration rate with respect to y_{kt}. Using (19), we can rewrite the gravity equation that could be brought to the data under the distributional assumptions introduced in Section 2 and when information is freely available as follows:

$$m_{jkt} = \exp\left[d_{kt} + d_{jt} + d_{jk} + \theta \ln(s_{jkt} + 1) + \ln(\zeta_{jkt})\right] \tag{20}$$

The inclusion of destination-time dummies d_{kt} would fail to absorb the effect of the time-varying attractiveness of destination k, as this effect would be heterogeneous across origins characterized by a different return from the acquisition of information when $\lambda_A > 0$. More specifically, rational inattention means that bilateral flows from countries where migrants have a higher return from acquiring information should be more responsive to variations in the attractiveness of country k. Thus, we can extend the specification of the gravity equation in (20) as follows:

$$m_{jkt} = \exp\left[d_{kt} + d_{jt} + d_{jk} + \beta\left[\ln(y_{kt}) \times \pi(r)_{jt}\right] + \theta \ln(s_{jkt} + 1) + \ln(\zeta_{jkt})\right] \tag{21}$$

where $\ln(y_{kt})$ is the logarithm of GDP per capita in destination k in year t,[23] and $\pi(r)_{jt}$, with $r = \{5, 10, 15, 20\}$, is the empirical counterpart of the return from information acquisition for migrants from the origin j leaving in the five-year period starting in t. Given that $\pi(r)_{jt} = w(r)_{jt} - c(\lambda_A)$, where $c(\lambda_A)$ is an unknown constant, the structure of fixed effects implies that (21) is equivalent to:[24]

$$m_{jkt} = \exp\left[d_{kt} + d_{jt} + d_{jk} + \beta\left[\ln(y_{kt}) \times w(r)_{jt}\right] + \theta \ln(s_{jkt} + 1) + \ln(\zeta_{jkt})\right] \tag{22}$$

---

[22] The data on the bilateral stock s_{jkt} comes from Özden et al. (2011) between 1960 and 2000, with interpolated values in between census years, and from United Nations Population Division (2015a) since 2005; the average and standard deviation of $\ln(s_{jkt} + 1)$ over our sample of 263,008 observations stand at 2.25 and 2.95 respectively.

[23] We use GDP per capita in 2010 USD from World Bank (2018); the average and standard deviation of $\ln(y_{kt})$ over our sample stand at 8.24 and 1.53 respectively.

[24] The same equivalence would still hold even if $\pi(r)_{jt} = w(r)_{jt} - c(\lambda_{At})$, i.e., if the marginal cost of acquiring information was varying over time.

as the destination-time dummies d_{kt} absorb the effect of $-\beta[\ln(y_{kt}) \times c(\lambda_A)]$, given that this part of the interaction term varies only at the destination-time level. The specification that we can bring to the data is thus given by (22): if migrants are rationally inattentive, then we should have that $\widehat{\beta} > 0$. Since we have a large share of zeros (61.2 percent) in our dependent variable m_{jkt}, we estimate (22) using a Poisson pseudo-maximum-likelihood estimator, following Santos Silva and Tenreyro (2006). More precisely, we employ the Stata command `ppmlhdfe` developed by Correia et al. (2019a,b), which allows handling in a computationally efficient way the large number of fixed effects in (22).

Dyadic dummies d_{jk} in (22) allow us to purge the estimate of the coefficient β of our interaction term from the influence of time-invariant determinants of dyadic migration costs that are hard to measure on a systematic basis, such as cultural distance (Spolaore and Wacziarg, 2016) or linguistic proximity between countries that do not share the same language (Adserà and Pytliková, 2015). For instance, an origin country j that is linguistically close to many destinations will typically have a lower share of its past flows directed to the main destination, i.e., higher value of $w(r)_{jt}$ and of the interaction term. The inclusion of d_{jk} thus prevents unobserved dyadic determinants of migration flows from biasing $\widehat{\beta}$. Origin-year d_{jt} dummies control for the influence exerted on m_{jkt} by $\ln\left(\sum_{l\in A} e^{v_{jlt}}\right)$ and $\ln(n_{jt})$ in (17), while destination-year d_{kt} dummies in (22) control for the time-varying attractiveness of the destination k.

4.2 Main results

Table 3 reports the results obtained when bringing the gravity equation described in (22) to the data. Each data column corresponds to one of the four variants of the empirical counterpart for the value of information $w(r)_{jt}$, with $r = \{5, 10, 15, 20\}$, for the origin country j in the five-year period starting in year t.

Table 3: Baseline results on the value of information

	Dependent variable: m_{jkt}			
	(1)	(2)	(3)	(4)
Value of r	5	10	15	20
$\ln(y_{kt}) \times w(r)_{jt}$	0.119***	0.145***	0.144***	0.161***
	(0.036)	(0.040)	(0.045)	(0.055)
$\ln(s_{jkt} + 1)$	0.193***	0.192***	0.195***	0.196***
	(0.023)	(0.023)	(0.023)	(0.023)
Observations	220,627	223,469	224,612	224,743
Pseudo-R^2	0.963	0.963	0.963	0.963
$w(r)_{jt}$ (mean)	0.866	0.925	0.953	0.968
$w(r)_{jt}$ (s.d.)	0.532	0.523	0.518	0.515
d_{jt}, d_{kt} and d_{jk}	Yes	Yes	Yes	Yes

Notes: ***, **, and * denote significance at the 1, 5, and 10 percent levels, respectively. Clustered standard errors at the origin-time level are reported in parentheses. The value of r denotes the number of years up to t that have been used to measure $w(r)_{jt}$. All regressions have been estimated with PPML using the Stata command `ppmlhdfe`.

Source: Authors' elaboration on Abel (2018), World Bank (2018), Özden et al. (2011) and United Nations Population Division (2015a).

The estimates reveal that the coefficient $\widehat{\beta}$ of the interaction between GDP per capita at destination and the time-varying origin-specific value of information is always positive and significant at the 1 percent confidence level, consistently with the empirical relevance of rational inattention in migration decisions.[25] A one standard deviation increase in the value of $w(r)_{jt}$ is associated with an increase in the elasticity of the bilateral migration rate with respect to GDP per capita at destination ranging between 0.063, in column (1), to 0.083, in column (4). Going back to the example of China and Mexico that we introduced in Section 3.2, the estimates in Table 3 entail that the elasticity for migration from China to any destination between 1995 and 2000 was 0.160-0.216 higher than the corresponding elasticity for migration from Mexico over the same time period. Similarly, the estimates also entail a substantial variability over time for a given origin; for instance, the elasticity of migration out

---

[25] Our analysis is fully robust to using gender-specific bilateral migration flows from Abel (2018); results are available from the Authors upon request.

of Ecuador increased by 0.069-0.093 between the early 1980s and the early 2000s,[26] following a substantial diversification of the main destinations for Ecuadorian migrants (Bertoli et al., 2011).

The lower responsiveness of bilateral migration flows originating from countries where migrants have stronger priors emerging from Table 3 could reflect the fact that a lower value of information induces migrants (*i*) to invest less in information acquisition across all destinations, or (*ii*) to selectively disregard some destinations that are *a priori* very unlikely to represent their preferred destinations, or (*iii*) both. The analysis of the independent consumer problem by Caplin et al. (2019), where a consumer has to select one good from a choice set, with the utilities associated to the various alternatives being independent, reveals that the optimal strategy of the consumer is to acquire information only about alternatives that are characterized by a probability to maximize consumer's utility that is above an endogenous threshold. Let $d^{\text{zero}}(5)_{jkt}$ be a dummy signaling a zero migration from j to k in the five years up to year t. We have that 57.9 percent of the observations in our sample correspond to origin-destination dyads with a zero flow in the recent past, and the migration flows for these dyads could be less sensitive to variations in economic conditions at destination, as migrants from j could exclude destination k from their (time-varying) consideration sets when $d^{\text{zero}}(5)_{jkt} = 1$.

---

[26] The value of $w(5)_{\text{ECU1980}}$ stood at 0.168, increasing to $w(5)_{\text{ECU2000}} = 0.744$.

Table 4: Zero past flows reduce responsiveness of flows to GDP per capita at destination

	Dependent variable: m_{jkt}		
	(1)	(2)	(3)
Value of r	5	5	5
$\ln(y_{kt}) \times d^{\text{zero}}(5)_{jkt}$	-0.030***	-0.028***	
	(0.007)	(0.007)	
$\ln(y_{kt}) \times w(r)_{jt}$		0.117***	0.119***
		(0.036)	(0.036)
$\ln(s_{jkt} + 1)$	0.192***	0.185***	0.193***
	(0.025)	(0.023)	(0.023)
Observations	220,627	220,627	220,627
Pseudo-R^2	0.963	0.964	0.963
$w(r)_{jt}$ (mean)	0.866	0.866	0.866
$w(r)_{jt}$ (s.d.)	0.532	0.532	0.532
d_{jt}, d_{kt} and d_{jk}	Yes	Yes	Yes

Notes: ***, **, and * denote significance at the 1, 5, and 10 percent levels, respectively. Clustered standard errors at the origin-time level are reported in parentheses. The value of r denotes the number of years up to t that have been used to measure $w(r)_{jt}$. $d^{\text{zero}}(5)_{jkt}$ is a dummy equal to 1 if $m_{jkt-5} = 0$, and 0 otherwise. All regressions have been estimated with PPML using the command `ppmlhdfe`.

Source: Authors' elaboration on Abel (2018), World Bank (2018), Özden et al. (2011) and United Nations Population Division (2015a).

Table 4 confirms that this is indeed the case: the elasticity with respect to GDP per capita at destination is 0.030 points lower for origin-destination dyads characterized by zero flows over the previous five years. However, this does not explain the role played by the value of information in our baseline results, as our coefficient of interest is only marginally reduced when introducing the additional interaction between $d^{\text{zero}}(5)_{jkt}$ and $\ln(y_{kt})$, as a comparison of the second and of the third data column in Table 4 reveals.[27]

---

[27] This also applies when using data over the previous 10, 15 or 20 years to identify origin-destination pairs with past zero flows; results are available from the Authors upon request.

4.3 Threats to our interpretation

The interpretation that we have provided of the estimates in Table 3 is threatened by a possible incorrect specification of location-specific utility in (18). More precisely, migration decisions could be subject to binding liquidity constraints, which could influence migrants' ability to respond to variations in economic conditions at destination even though they are able to costlessly observe them. Furthermore, location-specific utility might not be additively separable in y_{kt} and in τ_{jkt}, so that the semi-elasticity of v_{jkt} with respect to y_{kt} could be a function of the determinants of dyadic migration costs τ_{jkt}, e.g., the marginal utility of income might be a function of dyadic migration costs, or it might depend on migrants' individual characteristics such as education. Similarly, the specification of the gravity equation in (22) is consistent with a time-varying cost of information acquisition $c(\lambda_A)$, while it maintains that this does not vary across origins, but the cost of information acquisition could vary at the origin or at the dyadic level, possibly biasing our estimate of β.

We provide evidence here that our empirical evidence is robust to a more flexible specification of the deterministic component of utility that allows the responsiveness of bilateral migration flows with respect to variations in economic conditions at destination to be heterogeneous across origins or across origin-destination pairs.

4.3.1 Liquidity constraints

Figure 3 in Section 3.2 above revealed that the empirical counterpart $w(5)_{jt}$ for the value of information is higher in some geographical areas where most high-income countries are concentrated. If we rely on the classification by income groups from the World Bank, we have that the average value of $w(5)_{jt}$ stands at 1.088 for the high-income origin countries, while it is equal to 0.805 for the other origin countries in our sample.[28] Migration decisions can be subject to binding liquidity constraints (see, for instance, Clemens, 2014, Angelucci, 2015, Bazzi, 2017 and Dao et al., 2018), which entail that the set of affordable destinations is smaller than the choice set (Marchal and Naiditch, 2019), and hence this pattern in the data poses a threat to our interpretation of the results in Table 3. Migrants from lower-income countries might not value information less, but (because of liquidity constraints)

---

[28] The classification by the World Bank is available on an yearly basis since 1989; we use the classification for year t since 1990, and the earliest available classification for previous years for each origin. Source: datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups (last accessed on January 22, 2019).

they might be less able to react to variations in economic conditions at destination and their past distribution could be more concentrated in the main (affordable) destination. Thus, the positive and significant coefficient for the interaction term in Table 3 might be driven by liquidity constraints rather than by rational inattention.

Table 5: Heterogeneity by income group

	Dependent variable: m_{jkt}			
	(1)	(2)	(3)	(4)
Value of r	5	10	15	20
$\ln(y_{kt}) \times w(r)_{jt}$	0.119***	0.140***	0.139***	0.159***
	(0.036)	(0.040)	(0.045)	(0.056)
$\ln(y_{kt}) \times d_{jt}^{\text{low}}$	-0.170*	-0.145	-0.145	-0.131
	(0.098)	(0.098)	(0.102)	(0.104)
$\ln(y_{kt}) \times d_{jt}^{\text{l. middle}}$	-0.225**	-0.201**	-0.213**	-0.208**
	(0.092)	(0.091)	(0.096)	(0.097)
$\ln(y_{kt}) \times d_{jt}^{\text{u. middle}}$	-0.043	-0.032	-0.049	-0.040
	(0.085)	(0.085)	(0.090)	(0.090)
$\ln(s_{jkt}+1)$	0.201***	0.200***	0.203***	0.204***
	(0.023)	(0.023)	(0.023)	(0.023)
Observations	216,027	218,869	220,012	220,143
Pseudo-R^2	0.964	0.964	0.963	0.963
$w(r)_{jt}$ (mean)	0.867	0.931	0.965	0.981
$w(r)_{jt}$ (s.d.)	0.527	0.519	0.515	0.511
d_{jt}, d_{kt} and d_{jk}	Yes	Yes	Yes	Yes

Notes: ***, **, and * denote significance at the 1, 5, and 10 percent levels, respectively. Clustered standard errors at the origin-time level are reported in parentheses. d_{jt}^{low}, $d_{jt}^{\text{l. middle}}$, and $d_{jt}^{\text{u. middle}}$ are dummies taking the value of 1 if the origin j is classified by the World Bank as a low, lower middle-income or upper middle-income country respectively in year t, and 0 otherwise. The value of r denotes the number of years up to t that have been used to measure $w(r)_{jt}$. All regressions have been estimated with PPML using the Stata command `ppmlhdfe`.

Source: Authors' elaboration on Abel (2018), World Bank (2018), Özden et al. (2011) and United Nations Population Division (2015a).

We thus bring to the data an extended version of the gravity equation in (21), where we allow for a heterogeneous effect of $\ln(y_{kt})$ across groups of origins characterized by a different level of income, differentiating between low-income, lower middle income, upper middle income and high-income countries, with the latter representing the omitted category. The estimates in Table 5 reveal that the elasticity of the migration rate with respect to y_{kt} is significantly higher (or at least not lower) for origins classified as high-income countries in year t, consistently with the idea that liquidity constraints can reduce the responsiveness of migration flows to the time-varying attractiveness of a destination country. However, this does not influence either the size nor the significance of the coefficient for our interaction effect, thus dismissing the concern that the values of $\widehat{\beta}$ in Table 3 were picking up a spurious correlation between $w(r)_{jt}$ and the income group to which the origin j belonged in year t.[29]

4.3.2 More flexible responsiveness to economic conditions at destination

Do the results presented in Table 3 survive once we allow for a more general functional form of the deterministic component of utility v_{jkt}, or for differences across origins in the cost of acquiring information on the attractiveness of the various destinations? For instance, one could plausibly imagine that migrants from countries with larger past migration flows, with stronger networks at destination or facing lower moving costs could more easily acquire information on the attractiveness of the alternative destinations. This could explain our results if origin countries that are characterized by a higher value of information also face a lower cost of acquiring information.[30] We address this relevant empirical concern introducing an additional interaction term, between $\ln(y_{kt})$ and the logarithm of the total emigration rate for the origin j in the r years up to year t, with r taking the same value that is used to measure the value of information $w(r)_{jt}$. The estimated coefficient for this additional interaction term is always positive, and significant in three out of the four data columns in Table 6, in line with the idea that larger past migration flows reduce the cost of acquiring information on the attractiveness of the alternative destinations.[31] However, the inclusion of the additional

---

[29] We obtain similar results when considering a time-invariant income classification of the origin j, or when introducing an interaction between $\ln(y_{kt})$ and $\ln(y_{jt})$; results are available from the Authors upon request.

[30] Notice that the theoretical expression for the value of information in (14), as well as its empirical counterparts, are insensitive to the scale of past migration flows, as they only depend on the distribution of migrants across destinations.

[31] An alternative, but not mutually exclusive, explanation is that past migrants help relaxing liquidity constraints at origin through the remittances that they send back to their origin countries, thus increasing, as suggested by Table 5, the responsiveness of bilateral migration flows with respect to varying economic conditions at destination.

interaction term only marginally influences the size of the estimated value of $\widehat{\beta}$, and it leaves its significance unchanged. The estimated coefficients for the interaction between economic conditions at destination and value of information at origin range between 0.119 and 0.161 in Table 3, and between 0.096 and 0.168 in Table 6.

Table 6: Interaction with the past emigration rate

	Dependent variable: m_{jkt}			
	(1)	(2)	(3)	(4)
Value of r	5	10	15	20
$\ln(y_{kt}) \times w(5)_{jt}$	0.096^{***}	0.137^{***}	0.132^{***}	0.168^{***}
	(0.032)	(0.038)	(0.041)	(0.053)
$\ln(y_{kt}) \times \ln[\text{emigration rate}(r)_{jt}]$	0.020^{***}	0.014	0.042^{***}	0.066^{***}
	(0.007)	(0.019)	(0.013)	(0.017)
$\ln(s_{jkt} + 1)$	0.205^{***}	0.205^{***}	0.213^{***}	0.213^{***}
	(0.023)	(0.022)	(0.022)	(0.022)
Observations	213,545	216,031	216,805	216,570
Pseudo-R^2	0.964	0.964	0.964	0.964
$w(r)_{jt}$ (mean)	0.869	0.933	0.966	0.982
$w(r)_{jt}$ (s.d.)	0.527	0.519	0.516	0.513
d_{jt}, d_{kt} and d_{jk}	Yes	Yes	Yes	Yes

Notes: ***, **, and * denote significance at the 1, 5, and 10 percent levels, respectively. Clustered standard errors at the origin-time level are reported in parentheses. The value of r denotes the number of years up to t that have been used to measure $w(r)_{jt}$. $\ln[\text{emigration rate}(r)_{jt}]$ is the total emigration rate in the r years up to t for the origin country j. All regressions have been estimated with PPML using the Stata command `ppmlhdfe`.

Source: Authors' elaboration on Abel (2018), World Bank (2018), Özden et al. (2011) and United Nations Population Division (2015a).

Similarly, our empirical evidence is robust when interacting $\ln(y_{kt})$ with the (logarithm) of the size of the network of j-born migrants residing in destination k in year t, as shown in Table 7. Interestingly, the coefficient of this additional interaction term is negative and significant, suggesting that migration flows directed to destinations with larger diasporas from a given origin are less responsive to the varying attractiveness of various destinations. This might reflect the relevance of flows related to family reunification provisions, which are likely to be less responsive to business cycle conditions at destination.

Table 7: Heterogeneity with respect to the size of bilateral networks

	Dependent variable: m_{jkt}			
	(1)	(2)	(3)	(4)
Value of r	5	10	15	20
$\ln(y_{kt}) \times w(r)_{jt}$	0.106***	0.125***	0.120***	0.129***
	(0.030)	(0.034)	(0.038)	(0.046)
$\ln(y_{kt}) \times \ln(s_{jkt}+1)$	-0.070***	-0.066***	-0.066***	-0.066***
	(0.014)	(0.014)	(0.014)	(0.014)
$\ln(s_{jkt}+1)$	0.845***	0.802***	0.808***	0.807***
	(0.144)	(0.141)	(0.143)	(0.141)
Observations	220,627	223,469	224,612	224,743
Pseudo-R^2	0.964	0.964	0.964	0.963
$w(r)_{jt}$ (mean)	0.866	0.925	0.953	0.968
$w(r)_{jt}$ (s.d.)	0.532	0.523	0.518	0.515
d_{jt}, d_{kt} and d_{jk}	Yes	Yes	Yes	Yes

Notes: ***, **, and * denote significance at the 1, 5, and 10 percent levels, respectively. Clustered standard errors at the origin-time level are reported in parentheses. The value of r denotes the number of years up to t that have been used to measure $w(r)_{jt}$. All regressions have been estimated with PPML using the Stata command `ppmlhdfe`.

Source: Authors' elaboration on Abel (2018), World Bank (2018), Özden et al. (2011) and United Nations Population Division (2015a).

Table 8 similarly extends the gravity equation in (22) by introducing (either separately or jointly) interactions between the canonical dyadic controls from Mayer and Zignago (2011) and $\ln(y_{kt})$. The estimates reveal that origin-destination pairs with lower dyadic migration costs, e.g., contiguous countries, countries sharing a common language, countries with a past common colonial history and countries characterized by a lower geodesic distances, are characterized by a greater responsiveness of bilateral migration flows with respect to varying economic conditions at destination. However, this does not influence the estimated coefficient for $\ln(y_{kt}) \times w(5)_{jt}$, which ranges between 0.121 and 0.125, perfectly in line with $\widehat{\beta} = 0.119$ from the first data column in Table 3.[32,33]

---

[32] Similar evidence is obtained when modifying the value of r from 5 to either 10, 15 or 20; results are available from the Authors upon request.

[33] Additional results, which are available from the Authors upon request, reveal that our empirical evidence is also robust to the inclusion of interactions between y_{kt} and various measures of cultural and linguistic

Table 8: Heterogeneity with respect to dyadic determinants of migration costs

	Dependent variable: m_{jkt}				
	(1)	(2)	(3)	(4)	(5)
Value of r	5	5	5	5	5
$\ln(y_{kt}) \times w(r)_{jt}$	0.124***	0.121***	0.122***	0.124***	0.125***
	(0.035)	(0.036)	(0.036)	(0.034)	(0.035)
$\ln(y_{kt}) \times \text{Contiguity}_{jk}$	0.566***				0.288**
	(0.127)				(0.145)
$\ln(y_{kt}) \times \text{Common language}_{jk}$		0.258*			0.141
		(0.132)			(0.146)
$\ln(y_{kt}) \times \text{Colony}_{jk}$			-0.108		-0.175
			(0.159)		(0.183)
$\ln(y_{kt}) \times \ln(\text{distance}_{jk})$				-0.286***	-0.205***
				(0.052)	(0.056)
$\ln(s_{jkt} + 1)$	0.202***	0.200***	0.199***	0.201***	0.200***
	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)
Observations	214,838	214,838	214,838	214,838	214,838
Pseudo-R^2	0.964	0.964	0.964	0.964	0.964
$w(r)_{jt}$ (mean)	0.867	0.867	0.867	0.867	0.867
$w(r)_{jt}$ (s.d.)	0.531	0.531	0.531	0.531	0.531
d_{jt}, d_{kt} and d_{jk}	Yes	Yes	Yes	Yes	Yes

Notes: ***, **, and * denote significance at the 1, 5, and 10 percent levels, respectively. Clustered standard errors at the origin-time level are reported in parentheses. All regressions have been estimated with PPML using the Stata command `ppmlhdfe`.

Source: Authors' elaboration on Abel (2018), World Bank (2018), Mayer and Zignago (2011), Özden et al. (2011) and United Nations Population Division (2015a).

The stability of the coefficient $\widehat{\beta}$ for our main interaction term when we allow for the elasticity to vary across groups of origins or across origin-destination pairs is also reassuring with respect to the concern that the value of information might be picking up differences across origins in the composition of international migration flows that are associated with a differential responsiveness to economic conditions at destination. For instance, tertiary educated migrants might react differently to changing economic conditions at destination, but the inclusion of additional interactions of $\ln(y_{kt})$ with main origin-specific, i.e., income,

---

proximity between the origin j and the destination k from Spolaore and Wacziarg (2016) and Adserà and Pytliková (2015).

or bilateral, e.g., networks, correlates of the education composition of migration flows (see Beine et al., 2011) allows, at least partly, to downplay this concern.

5 Concluding remarks

The insights obtained from applying the theory of rational inattention to discrete choice models are relevant to enhance our understanding of the determinants of international migration flows. We have proposed a theoretical innovation that introduces in an otherwise standard random utility maximization model a cost of acquiring information on the attractiveness of alternative destinations. The proposed extension to the analysis of the location-decision problem that migrants face entails that the strength of the priors on the identity of the utility-maximizing destination determines the extent to which migrants will rationally invest in information acquisition. This, in turn, implies that location choices will be less sensitive to variations in the attractiveness of alternative destinations when migrants have stronger priors. The econometric evidence that we provide using data on bilateral migration flows between 1980 and 2015 is consistent with this testable prediction, and robust to alternative explanations of the observed differences across origin countries in the elasticity of the bilateral migration rate with respect to income per capita at destination.

References

Abel, G. J. (2018): "Estimates of Global Bilateral Migration Flows by Gender between 1960 and 2015," *International Migration Review*, 52, 809–852.

Abel, G. J. and N. Sander (2014): "Quantifying Global International Migration Flows," *Science*, 343, 1520–1522.

Adserà, A. and M. Pytliková (2015): "The role of language in shaping international migration," *Economic Journal*, 125, F49–F81.

Anderson, S., A. de Palma, and J.-F. Thisse (1992): *Discrete Choice Theory of Product Differentiation*, Cambridge, MA: The MIT Press.

Angelucci, M. (2015): "Migration and financial constraints: Evidence from Mexico," *Review of Economics and Statistics*, 97, 224–228.

Artuç, E. and Ç. Özden (2018): "Transit migration: All roads lead to America," *Economic Journal*, 128, F306–F334.

Batista, C. and D. McKenzie (2018): "Testing Classic Theories of Migration in the Lab," Mimeo.

Bazzi, S. (2017): "Wealth Heterogeneity and the Income Elasticity of Migration," *American Economic Journal: Applied Economics*, 9, 219–255.

Beine, M., F. Docquier, and Ç. Özden (2011): "Diasporas," *Journal of Development Economics*, 95, 30–41.

Ben-Akiwa, M. and S. Lerman (1979): "Disaggregate Travel and Mobility Choice Models and Measures of Accessibility," in *Behavioral Travel Modelling*, ed. by D. Hensher and P. Stopher, London: Croom Helm, 654–679.

Bertoli, S. (2010): "The informational structure of migration decision and migrants' self-selection," *Economics Letters*, 108, 89–92.

Bertoli, S., H. Brücker, and J. Fernández-Huertas Moraga (2016): "The European crisis and migration to Germany," *Regional Science and Urban Economics*, 60, 61–72.

Bertoli, S. and J. Fernández-Huertas Moraga (2013): "Multilateral Resistance to Migration," *Journal of Development Economics*, 102, 79–100.

Bertoli, S., J. Fernández-Huertas Moraga, and F. Ortega (2011): "Immigration Policies and the Ecuadorian Exodus," *World Bank Economic Review*, 25, 57–76.

Borjas, G. J. (1987): "Self-Selection and the Earnings of Immigrants," *American Economic Review*, 77, 531–553.

Caplin, A., M. Dean, and J. Leahy (2017): "Rationally Inattentive Behavior: Characterizing and Generalizing Shannon Entropy," Working Paper 23652, National Bureau of Economic Research.

——— (2019): "Rational Inattention, Optimal Consideration Sets, and Stochastic Choice," *Review of Economic Studies*, 86, 1061–1094.

Caplin, A., J. Leahy, and F. Matějka (2016): "Rational Inattention and Inference from Market Share Data," mimeo.

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

Clemens, M. (2014): "Does development reduce migration?" in *International Handbook on Migration and Economic Development*, ed. by R. Lucas, Cheltenham: Edward Elgar Publishing, 152–185.

Clemens, M. A., C. Montenegro, and L. Pritchett (2018): "The Place Premium: Bounding the Price Equivalent of Migration Barriers," *Review of Economics and Statistics*, https://doi.org/10.1162/rest_a_00776.

Correia, S., P. Guimarães, and T. Zylkin (2019a): "ppmlhdfe: Fast Poisson Estimation with High-Dimensional Fixed Effects," ArXiv:1903.01690.

——— (2019b): "Verifying the existence of maximum likelihood estimates for generalized linear models," ArXiv:1903.01633.

Cover, T. M. and J. A. Thomas (1991): *Elements of Information Theory*, New York: Wiley.

Dao, T. H., F. Docquier, C. Parsons, and G. Peri (2018): "Migration and development: Dissecting the anatomy of the mobility transition," *Journal of Development Economics*, 132, 88–101.

Dasgupta, K. and J. Mondria (2018): "Inattentive Importers," *Journal of International Economics*, 112, 150–165.

Eaton, J. and S. Kortum (2002): "Technology, geography, and trade," *Econometrica*, 70, 1741–1779.

Fosgerau, M., E. Melo, A. de Palma, and M. Shum (2018): "Discrete Choice and Rational Inattention: a General Equivalence Result," Mimeo, http://dx.doi.org/10.2139/ssrn.2889048.

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

Maćkowiak, B., F. Matějka, and M. Wiederholt (2018): "Rational Inattention, a Disciplined Behavioral Model," CEPR Discussion Papers 13243, London, CEPR.

Marchal, L. and C. Naiditch (2019): "How Borrowing Constraints Hinder Migration: Theoretical Insights from a RUM Model," *Scandinavian Journal of Economics*, https://doi.org/10.1111/sjoe.12355.

Matějka, F. and A. McKay (2015): "Rational Inattention to Discrete Choices: A New Foundation for the Multinomial Logit Model," *American Economic Review*, 105, 272–298.

Mayer, T. and S. Zignago (2011): "Notes on CEPII's distances measures: The GeoDist database," CEPII Working Paper No. 2011-25, Paris.

McFadden, D. (1974): "Conditional logit analysis of qualitative choice behavior," in *Frontiers in Econometrics*, ed. by P. Zarembka, New York: Academic Press, 105–142.

——— (1978): "Modeling the Choice of Residential Location," in *Spatial interaction theory and planning models*, ed. by A. Karlqvist, L. Lundqvist, F. Snickars, and J. Weibull, Amsterdam: North-Holland, 75–96.

McKenzie, D., J. Gibson, and S. Stillman (2013): "A land of milk and honey with streets paved with gold: Do emigrants have over-optimistic expectations about incomes abroad?" *Journal of Development Economics*, 102, 116–127.

McKenzie, D., S. Stillman, and J. Gibson (2010): "How important is selection? Experimental vs. non-experimental measures of the income gains from migration," *Journal of the European Economic Association*, 8, 913–945.

Nelson, P. (1970): "Information and Consumer Behavior," *Journal of Political Economy*, 78, 311–329.

Ortega, F. and G. Peri (2013): "The Role of Income and Immigration Policies in Attracting International Migrants," *Migration Studies*, 1, 47–74.

Özden, Ç., C. R. Parsons, M. Schiff, and T. L. Walmsley (2011): "Where on Earth is Everybody? The Evolution of Global Bilateral Migration 1960-2000," *World Bank Economic Review*, 25, 12–56.

Roy, A. D. (1951): "Some thoughts on the distribution of earnings," *Oxford Economic Papers*, 3, 135–146.

Santos Silva, J. M. C. and S. Tenreyro (2006): "The Log of Gravity," *Review of Economics and Statistics*, 88, 641–658.

Shannon, C. E. (1948): "A mathematical theory of communication," *Bell System Technical Journal*, 27, 379–423.

Simon, H. A. (1959): "Theories of decision-making in economics and behavioral science," *American Economic Review*, 49, 253–283.

Sims, C. A. (1998): "Stickiness," *Carnegie-Rochester Conference Series on Public Policy*, 49, 317–356.

——— (2003): "Implications of rational inattention," *Journal of Monetary Economics*, 50, 665–690.

Small, K. A. and H. S. Rosen (1981): "Applied welfare economics with discrete choice models," *Econometrica*, 49, 105–130.

Spolaore, E. and R. Wacziarg (2016): "Ancestry, language and culture," in *The Palgrave Handbook of Economics and Language*, ed. by V. Ginsburgh and S. Weber, London: Palgrave McMillan, 174–211.

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

United Nations Population Division (2015a): *Trends in International Migrant Stock: The 2015 Revision*, New York, NY: United Nations, Department of Economic and Social Affairs.

——— (2015b): *World Population Prospects: The 2015 Revision*, New York, NY: United Nations, Department of Economic and Social Affairs.

World Bank (2018): *World Development Indicators*, Washington DC: the World Bank.

A Appendix

A.1 The solution to the location-decision problem

Let $\boldsymbol{u}_{ij} = \boldsymbol{v}_{ij} + \boldsymbol{\epsilon}_{ij} = (u_{ij1}, u_{ij2}, ..., u_{ijN})'$ represent a $N \times 1$ column vector stacking the utility that migrant i from country j associates to each one of the N potential destination in the choice set A. The solution to the location-decision problem described in Section 2 when $\lambda_k = \lambda > 0$, $\forall k \in A$, is given by (Matějka and McKay, 2015):

$$p_{ijk}(\boldsymbol{u}_{ij}) = \frac{p^0_{ijk} e^{u_{ij1}/\lambda}}{\sum_{l=1}^{N} p^0_{ijl} e^{u_{il1}/\lambda}} \tag{A.1}$$

where:

$$p^0_{ijk} \equiv \mathbb{E}_{\boldsymbol{u}_{ij}}[p_{ijk}(\boldsymbol{u}_{ij})] = \int_{\boldsymbol{u}_{ij}} p_{ijk}(\boldsymbol{u}_{ij}) g(\boldsymbol{u}_{ij}) \, \mathrm{d}\boldsymbol{u}_{ij} \tag{A.2}$$

where $g(\boldsymbol{u}_{ij})$ represents the prior held by the migrant i on the distribution of location-specific utility, and we assume that $g(\boldsymbol{u}_{ij}) = f(\boldsymbol{u}_{ij})$. Notice that the functional form of the conditional choice probabilities in (A.1) do not depend on the distributional assumptions on location-specific utility, which enter into (A.1) only via the unconditional probability p^0_{ijk} defined in (A.2). The collection of unconditional choice probabilities $p^0_{ij} = \{p^0_{ijk}\}_{k=1}^{N}$ is optimal if and only if (Caplin et al., 2019):

$$\int_{\boldsymbol{u}_{ij}} \frac{e^{u_{ij1}/\lambda}}{\sum_{l=1}^{N} p^0_{ijl} e^{u_{il1}/\lambda}} g(\boldsymbol{u}_{ij}) \, \mathrm{d}\boldsymbol{u}_{ij} \leq 1 \tag{A.3}$$

for all $k \in A$, with equality for any alternative k such that $p^0_{ijk} > 0$, and where conditional choice probabilities are defined by (A.1). Caplin et al. (2019) define the consideration $B(p^0_{ij})$ as the subset of A including all the alternatives that are chosen with a positive probability. The expected utility $V(\lambda)$ from the choice situation is given by:

$$V(\lambda) \equiv \max_{\{p^0_{ijk}\}_{k=1}^{N}} \int_{\boldsymbol{u}_{ij}} \lambda \log \left(\sum_{k=1}^{N} p^0_{ijk} e^{u_{ijk}/\lambda} \right) g(\boldsymbol{u}_{ij}) \, \mathrm{d}\boldsymbol{u}_{ij} \tag{A.4}$$

which is given by the difference from the gross expected utility $\mathbb{E}_{\boldsymbol{u}_{ij}}[\sum_{k=1}^{N} u_{ijk} p_{ijk}(\boldsymbol{u}_{ij})]$ and the cost of information acquisition $C(\lambda)$ that corresponds to the optimal choice probabilities. The latter is given by the product between the marginal cost of information acquisition λ

and the reduction in entropy:

$$C(\lambda) \equiv \lambda \left[-\sum_{k=1}^{N} p^0_{ijk} \log p^0_{ijk} + \int_{\boldsymbol{u}_{ij}} \left(\sum_{k=1}^{N} p_{ijk}(\boldsymbol{u}_{ij}) \log p_{ijk}(\boldsymbol{u}_{ij}) \right) g(\boldsymbol{u}_{ij})\, \mathrm{d}\boldsymbol{u}_{ij} \right] \tag{A.5}$$

where $p_{ijk}(\boldsymbol{u}_{ij})$ is defined by (A.1), and p^0_{ijk} is defined by (A.2) and satisfies (A.3). In general, no closed-form expression exists for the optimal conditional and unconditional choice probabilities

A.2 Identical deterministic component of utility

If we assume that all alternatives in the choice set are characterized by an identical deterministic component of utility, i.e., $v_{ijk} = v_{ij}$, $\forall k \in A$,[34], then we can write down a closed-form expression for (A.1) and (A.2), as we have that (Matějka and McKay, 2015; Caplin et al., 2019):

$$p_{ijk}(\boldsymbol{u}_{ij}) = \frac{e^{u_{ij1}/\lambda}}{\sum_{l=1}^{N} e^{u_{il1}/\lambda}} \tag{A.6}$$

and:

$$p^0_{ijk} = \frac{1}{N} \tag{A.7}$$

Notice that in such a case the value of acquiring full information $w_{ij} = V(0) - \lim_{\lambda \to \infty} V(\lambda)$ defined in (11) is simply equal to $\log N$. We can also rewrite the (optimal) cost of information acquisition in (A.5) can be simplified as follows:

$$C(\lambda) = \lambda \int_{\boldsymbol{u}_{ij}} \left[\sum_{i=1}^{N} \frac{u_{ijk}}{\lambda} p_{ijk}(\boldsymbol{u}_{ij}) - \log \left(\sum_{l=1}^{N} e^{u_{ijl}/\lambda} \right) \right] g(\boldsymbol{u}_{ij})\, \mathrm{d}\boldsymbol{u}_{ij} + \lambda \log N \tag{A.8}$$

Consider the two terms in the integrand function that appear in (A.9): the second term corresponds to the expected value from the choice situation when the alternative-specific utility follows an i.i.d. EVT-1 distribution, the deterministic component of utility v_{ijk} is equal to u_{ijk}/λ and information can be costlessly acquired. Fosgerau et al. (2018) demonstrate

---

[34] We can normalize v_{ij} to 0 without loss of generality.

that, with these distributional assumptions:

$$\log\left(\sum_{l=1}^{N} e^{v_{ijk}}\right) = \sum_{k=1}^{N} p_{ijk}\left[v_{ijk} + \mathbb{E}_{\boldsymbol{\epsilon}_{ij}}\left(\epsilon_{ijk}|u_{ijk} \geq u_{ijl}, l \in A\right)\right]$$

where p_{ijk} are the choice probabilities of a canonical conditional logit model. This entails that we always have that:

$$\sum_{i=1}^{N} v_{ijk}p_{ijk}(\boldsymbol{u}_{ij}) - \log\left(\sum_{l=1}^{N} e^{v_{ijk}}\right) = -\sum_{k=1}^{N} p_{ijk}\mathbb{E}_{\boldsymbol{\epsilon}_{ij}}\left(\epsilon_{ijk}|u_{ijk} \geq u_{ijl}, l \in A\right) < 0$$

This allows us to conclude that the integrand in (A.9) always negative, so that:

$$C(\lambda) \leq \lambda \log N = \lambda\left[V(0) - \lim_{\lambda\to\infty} V(\lambda)\right] = \lambda w_{ij} \qquad \text{(A.9)}$$

The actual investment in information acquisition $C(\lambda)$ that corresponds to the solution of the location-decision problem with *ex ante* identical alternatives is bounded from above by a term that is proportional to the expected value of acquiring full information, with the constant of proportionality being given by λ. When migrants' priors become weaker, i.e., N increases, the upper bound of $C(\lambda)$ increases, and migrants rationally decide to invest more in information acquisition.