Rational inattention and migration decisions*

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

[a] *Université Clermont Auvergne, CNRS, CERDI, IUF, IZA*†

[b] *Universidad Carlos III de Madrid*‡

[c] *IAB, Université Clermont Auvergne, CNRS, CERDI*§

Abstract

Acquiring information about destinations can be costly for migrants. We model information frictions in the rational inattention framework and obtain a closed-form expression for a migration gravity equation that we bring to the data. The model predicts that flows from countries with a higher cost of information or stronger priors are less responsive to variations in economic conditions in the various destinations, as migrants rationally get less information before deciding where to move. The econometric analysis reveals systematic heterogeneity in the pro-cyclical behavior of migration flows across origins that is consistent with the existence of information frictions.

Keywords: international migration; information; rational inattention; gravity equation.

JEL codes: F22; D81; D83.

*The Authors are grateful to the Editor Treb Allen and to two anonymous referees for their comments, and to Rabah Arezki, Erhan Artuç, Atanas Christev, Mark Dean, Vianney Dequiedt, Lionel Fontagné, Nicholas Hanley, Joël Machado, Thierry Mayer, David McKenzie, Çağlar Özden, Panu Poutvaara, Ariell Reshef, Mark Schaffer, Victor Stephane, Jérôme Valette, and to the participants at various conferences and seminars for their suggestions; 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 for providing valuable research assistance; Simone Bertoli and Lucas Guichard acknowledge 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); Jesús Fernández-Huertas Moraga acknowledges the financial support from the *Ministerio de Ciencia e Innovación* (Spain), grants PID2019-111095RB-I00 and MDM 2014-0431; 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

Human migration is portrayed as an investment decision that should be based on a comparison of the private returns for the migrant in each of the potential destinations (Sjaastad, 1962), but the key elements that lead to the choice of the preferred destination are unlikely to be readily available. The migrant needs first to gather information about the attractiveness of the various countries she could opt for. However, some of the seminal contributions to the modeling of the determinants of migration choice assume that uncertainty is fully (and costlessly) resolved before deciding where to migrate.[1] In particular, this is the case for the canonical micro-foundations of migration gravity equations that rely on discrete choice models *à la* McFadden (McFadden, 1974). In contrast, there is empirical evidence revealing that potential migrants can have inaccurate expectations on their earnings abroad (McKenzie et al., 2013) or about the costs and risks associated to migrating (Shrestha, 2020).

This suggests that the uncertainty surrounding the utility at destination might not be entirely resolved when a migrant has to come up with a decision, and the size of the remaining uncertainty could be endogenously determined. 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.

How can we enhance our understanding of the determinants of international migration flows if we take into account the uncertainty that migrants face, and the costly actions that they can take to narrow it down? We estimate a gravity equation whose specification is derived from the analysis of a location-decision problem with information frictions. We obtain

[1] 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 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).

a closed-form expression for optimal choice probabilities under suitable assumptions on the priors held by the migrants about the distribution of destination-specific utility, following Dasgupta and Mondria (2018).[2] The main testable implication of this model is that the responsiveness of bilateral migration flows with respect to variations in the attractiveness of alternative destinations is larger when migrants have a stronger incentive to acquire information before deciding where to move. We refer to this incentive as the value of information, which is related to the ratio between the variance of the prior distribution of destination-specific utility and the marginal cost of receiving signals about the actual attractiveness of the various alternatives in the choice set. The distribution of past migration flows across destinations can be used to infer the (unobserved) value of information, and we exploit this property to estimate the model.

We draw on data on bilateral migration flows between 1960 and 2015 from Abel (2018) to build an origin-specific and time-varying measure of the value of information for international migrants, which is inversely related to the share of cumulated past flows directed to the main destination.[3] We estimate a gravity equation where the destination-specific utility depends on an interaction between income per capita at destination and our empirical counterpart 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.[4] Our estimates imply that the elasticity of the bilateral migration rate with respect to income per capita for China is 0.182-0.241 higher than the corresponding elasticity for Mexico, which represents a paradigmatic case of migration flows concentrated in just one single destination, namely the United States. Our results are robust when we exclude the main origin-specific destination from the sample, so that they are are not driven by a lower procyclicality of the migration flows directed to just one destination but rather, as the theory predicts, to all foreign countries. Our results are inconsistent with the predictions stemming from a canonical random utility maximization model with unobserved heterogeneity, where the variance of the stochastic component of utility is origin-

[2] Dasgupta and Mondria (2018) have drawn on Matějka and McKay (2015) to extend the N-country Ricardian model of trade by Eaton and Kortum (2002), introducing costly acquisition of information on the prices of goods in different exporting countries.

[3] 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).

[4] Consistently with a theoretical result derived by Dasgupta and Mondria (2018), we obtain a non-significant coefficient for this interaction term when we measure the value of information using the past share of migrants in destinations other than the main one.

specific. This alternative full-information model would imply that the coefficient of our interaction term should have the opposite sign to the one that we obtain when estimating our gravity equation.

The econometric evidence that we provide is fully robust when we allow for additional heterogeneity in the coefficient of income at destination either across origins or at the dyadic level. Specifically, we let this coefficient vary also with the level of income of the migrant-sending country, with its past total emigration rate, and with dyadic correlates of migration costs, such as the size of migrant networks at destination, geographic, cultural or linguistic distance. This, in turn, implies that our results cannot be explained by a full-information model with a richer and more flexible specification of the deterministic component of utility, where the effect of income at destination depends in a multiplicative way on other variables, which might also be correlated with the past distribution of flows across destinations. Thus, the results of the estimation of our theory-based gravity equation suggest that variations in economic conditions in a given destination country influence more incoming migration flows from origins where migrants (rationally) invest more in information acquisition.

This paper is mainly related to two strands of literature, namely (*i*) the theoretical analyses of discrete choice models with costly information acquisition (Matějka and McKay, 2015; Caplin et al., 2019; Fosgerau et al., 2020; Steiner et al., 2017), and (*ii*) the analysis of the determinants of international migration flows through micro-founded specifications of the gravity equation (see, for instance, Mayda, 2010; Grogger and Hanson, 2011; Bertoli and Fernández-Huertas Moraga, 2013; Ortega and Peri, 2013).[5] With respect to (*i*), we make three distinct contributions to the literature on rational inattention. First, we prove that all alternatives are chosen with positive probability,[6] once we assume that utility is identically and independently distributed according to a conjugate of a Gumbel distribution (Cardell, 1997) around a destination-specific expected value.[7] Second, we show that the optimal total investment in information acquisition is negatively related to the expected utility associated to the alternative that is, *a priori*, most attractive, but that the migrant

[5] Batista and McKenzie (2018) have recently 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.

[6] This is a natural property in models of industrial organization, e.g., Brown and Jeon (2020), where profit-maximizing rules out prices that would bring the demand to zero, but needs to be demonstrated in settings in which the attractiveness of the various alternatives is not endogenously determined.

[7] "Determining the empirical content of the rational inattention model with nonexchangeable priors [...] is an active area of research" (Natenzon, 2019, p. 445), and our paper thus also contributes to develop the analysis of models where the priors about the distribution of utility are alternative-specific.

chooses to receive more informative (and hence costly) signals about the alternatives that are less likely to be selected. This latter theoretical result is reminiscent of the evidence about the redirection of attention towards less attractive options in the so-called lemon-dropping markets in Bartoš et al. (2016). Third, we provide evidence of the empirical relevance of rational inattention in discrete choice situations, complementing a strand of literature that is still mostly theoretical.[8] 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., 2019).

As far as (*ii*) is concerned, ours is the first paper bringing to the data a migration gravity equation derived from a model with information frictions, with Porcher (2019) being the only other paper we are aware of, in his case exploiting internal migration flows in Brazil. Furthermore, we make two main contributions. First, we demonstrate that an alternative micro-foundation of the migration gravity equation allows for uncovering and interpreting systematic heterogeneities across origins in the responsiveness of migration flows with respect to varying economic conditions in the various destination countries. Second, our analysis implies an additional reason why migration flows have an inertial character, over and above the positive externalities generated due to destination-specific migration networks (e.g., Munshi, 2003), as information frictions induce a more concentrated distribution of migrants across destinations.[9]

The rest of the paper is structured as follows: Section 2 introduces information frictions in a standard location-choice problem, solving it under suitable distributional assumptions, and deriving its testable implications; Section 3 briefly presents the main data sources, it describes how we bring the model to the data, and it presents basic descriptive statistics. Section 4 presents the results of the econometric analysis, and Section 5 concludes.

2 Theoretical model

Consider a migrant from the origin country j who has to select her preferred destination from a choice set A including N alternatives, i.e., foreign countries, so that we analyze the

[8] "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).

[9] Our results also reveal an additional dimension of interdependence between migration flows directed to different countries, beyond the strategic interactions in migration policies (Giordani and Ruta, 2013).

choice of the destination conditional upon migrating. Let $v_{jk} = w_k - c_{jk}$ denote the utility, or payoff, associated to alternative $k \in A$, and let $\boldsymbol{v}_j \in \mathbb{R}^N$ represent the vector of payoffs, which we will be referring to as the state of the world. We omit the origin subscript j to avoid cluttering the notation, but the distribution of payoffs (because of the dyadic migration costs c_{jk}) and all the other parameters of the model can be origin-specific.

We denote by $F(\boldsymbol{v})$ the belief held by the migrant on the distribution of the state of the world; we assume that $F(\boldsymbol{v})$ is differentiable, and we denote by $f(\boldsymbol{v})$ the probability density function. We define $\overline{v}_k \equiv \int_{\boldsymbol{v}} v_k f(\boldsymbol{v})\,\mathrm{d}\boldsymbol{v}$, and we assume that the expected value of the payoff is finite. Without loss of generality, we also assume that $\overline{v}_k \geq \overline{v}_h$, when $k < h$, $\forall k, h \in A$.

The migrant can obtain a signal $\boldsymbol{s} \in \mathbb{R}^N$ about the payoffs in the various alternatives in the choice set, choosing both where to focus her attention (e.g., some destinations might be completely disregarded), and how much information to acquire before deciding where to migrate. More precise signals, i.e., signals that convey more information about the state of the world, are more costly, and the cost of information acquisition is proportional to the mutual information between the signal $\boldsymbol{s}$ and the state $\boldsymbol{v}$. The parameter $\lambda > 0$ translates the reduction in the entropy of $\boldsymbol{v}$ (Shannon, 1948) induced by the chosen information acquisition strategy into the same metrics as the payoffs.[10] The migrant behaves as a Bayesian expected utility maximizer, selecting the alternative in A with the highest expected payoff given the posterior distribution of $\boldsymbol{v}$ that has been induced by the signal $\boldsymbol{s}$, i.e., $F(\boldsymbol{v}|\boldsymbol{s})$.

Letting $S_k \subseteq \mathbb{R}^N$ be the set of signals that induces the migrant to select $k \in A$, the probability of opting for alternative k under the state of the world $\boldsymbol{v}$ is given by:

$$\mathcal{P}_k(\boldsymbol{v}) \equiv \int_{\boldsymbol{s} \in S_k} F(\mathrm{d}\boldsymbol{s}|\boldsymbol{v})$$

A key property of this model is that the migrant is never going to acquire distinct signals that lead to the choice of the same alternative, as in this case costly information would be acquired but not acted upon. This implies that the mutual information between the state and the signal is the same as the mutual information between the state and the alternative. This fundamental result (see Lemma 1 in Matějka and McKay, 2015), coupled with the symmetry of mutual information, implies that we can cast the location-decision problem facing the migrant in terms of the selection of the conditional choice probabilities $\mathcal{P}_k(\boldsymbol{v})$, $\forall k \in A$. The

[10] This parameter is invariant across alternatives in the choice set; if λ was alternative-specific, conditional choice probabilities would no longer have the functional form derived by Matějka and McKay (2015).

location-decision problem that the migrant faces can thus be described as follows:[11]

$$\max_{\mathcal{P}=\{\mathcal{P}_a(\boldsymbol{v})\}_{a=1}^N} \sum_{a=1}^N \int_{\boldsymbol{v}} v_a \mathcal{P}_a(\boldsymbol{v}) f(\boldsymbol{v})\,\mathrm{d}\boldsymbol{v} - C(\mathcal{P}), \tag{1}$$

where:

$$C(\mathcal{P}) \equiv \sum_{a=1}^N C_a(\mathcal{P}),\ C_a(\mathcal{P}) = \lambda\left(-\mathcal{P}_a \ln \mathcal{P}_a + \int \mathcal{P}_a(\boldsymbol{v}) \ln \mathcal{P}_a(\boldsymbol{v}) f(\boldsymbol{v})\,\mathrm{d}\boldsymbol{v}\right), \tag{2}$$

with $\mathcal{P}_a \equiv \int_{\boldsymbol{v}} \mathcal{P}_a(\boldsymbol{v}) f(\boldsymbol{v})\,\mathrm{d}\boldsymbol{v}$, and subject to the constraints:

$$\mathcal{P}_a(\boldsymbol{v}) \geq 0,\ \forall a \in A, \forall \boldsymbol{v} \in \mathbb{R}^N,\ \sum_{a=1}^N \mathcal{P}_a(\boldsymbol{v}) = 1,\ \forall \boldsymbol{v} \in \mathbb{R}^N. \tag{3}$$

The location-decision problem described in (1)-(3) is characterized by the parameter $\lambda > 0$, and by the function $f(\boldsymbol{v})$ that denotes the distribution of the vector of payoffs.

2.1 Solution of the model

Matějka and McKay (2015) prove in Theorem 1 that the optimal conditional probability for $k \in B$ is given by:[12]

$$\mathcal{P}_k(\boldsymbol{v}) = \frac{\mathcal{P}_k e^{v_k/\lambda}}{\sum_{a\in B} \mathcal{P}_a e^{v_a/\lambda}} \tag{4}$$

where $\mathcal{P}_k \equiv \int_{\boldsymbol{v}} \mathcal{P}_k(\boldsymbol{v}) f(\boldsymbol{v})\,\mathrm{d}\boldsymbol{v}$. We denote by $B \subseteq A$ the consideration set (Caplin et al., 2019), i.e., the set of alternatives that are chosen with positive probability.

If we plug the expression for $\mathcal{P}_k(\boldsymbol{v})$ in (4) in the original maximization problem in (1), this can be expressed only in terms of the unconditional probabilities:[13]

$$\max_{\mathcal{P}_1,\ldots,\mathcal{P}_N} \int_{\boldsymbol{v}} \lambda \ln\left[\sum_{a\in B} \mathcal{P}_a e^{v_a/\lambda}\right] f(\boldsymbol{v})\,\mathrm{d}\boldsymbol{v} \tag{5}$$

The analytical challenges that are related to the solution of the model are that (i) we do

[11] In the expression for entropy, we adopt the convention that $0\ln(0) = 0$.

[12] See Fosgerau et al. (2020) on the relationship between the use of Shannon entropy to define $C(\mathcal{P})$ and the functional form of optimal conditional choice probabilities.

[13] This is Lemma 2 in Matějka and McKay (2015).

not know what is the composition of the set B,[14] and (ii) a closed-form expression for the integral in (5) does not, in general, exist.

2.1.1 Consideration set

With respect to point (i), the number of potential sets of alternatives that correspond to the solution of the maximization problem in (1) stands in general at $2^N - 1$. If $v_k = \overline{v}_k + \epsilon_k$ and ϵ_k is identically and independently distributed for all alternatives $k \in A$, then there are just N different subsets of A that can be the consideration set, and these are nested. This is implied by Theorem 2 in Caplin et al. (2019); when payoffs are independently distributed across alternatives, if $k \in B$, i.e., $\mathcal{P}_k > 0$, then $l \in B$ if:

$$\int_{-\infty}^{+\infty} e^{(\overline{v}_l+\epsilon_l)/\lambda} f(\epsilon_l)\,\mathrm{d}\epsilon_l \geq \int_{-\infty}^{+\infty} e^{(\overline{v}_l+\epsilon_k)/\lambda} f(\epsilon_k)\,\mathrm{d}\epsilon_k \tag{6}$$

When ϵ_k and ϵ_l are identically distributed, then the distribution of the payoff for alternative k is first-order stochastically dominated by the distribution of the payoff for alternative l, $\forall l < k$. Thus, if an alternative $k \in B$, then $l \in B$ for all alternatives $l = 1, ..., k-1$, and the consideration set can only be of the type $B_k = \{1, ..., k\}$, with $k = 1, ..., N$.

2.1.2 Solving for unconditional probabilities

As far as point (ii) is concerned, a closed-form solution for the unconditional probabilities can be obtained by assuming that the distribution of payoffs is the same across all alternatives, so that $\overline{v}_k = \overline{v}$, $\forall k \in A$, or by allowing for alternative-specific values of the expected payoff under suitable distributional assumptions. If payoffs are identically distributed for all alternatives, then the consideration set is $B_N = A$, and all alternatives are chosen with probability $1/N$.[15]

The second option is to introduce the same distributional assumptions as in Dasgupta and Mondria (2018), Brown and Jeon (2020) and Porcher (2019). We can thus assume that $v_k = \overline{v}_k + \epsilon_k$, where ϵ_k is identically and independently drawn according to a Cardell distribution $C(\lambda)$, with $\lambda \in (0, 1)$. The key property of this distribution, whose density is fully supported on the real line, is that it is the (unique) conjugate of the EVT-1 distribution: when η_k is EVT-1 and ϵ_k is an independent $C(\lambda)$ random variable, then $\epsilon_k + \lambda\eta_k$ follows an

[14] Caplin et al. (2019) derive necessary and sufficient conditions to have that $P_k > 0$.
[15] See Proposition 1 in Matějka and McKay (2015).

EVT-1 distribution (Cardell, 1997).[16,17]

With these distributional assumptions, once we fix λ we are also pinning down σ^2, but the ratio between the variance of the payoffs and the marginal cost of acquiring information can take any positive value when $\lambda \in (0,1)$, as $\sigma^2/\lambda = \frac{(1-\lambda^2)}{\lambda}(\pi^2/6)$. Thus, we can represent a location-decision problem with an arbitrary quantity associated to the ratio between the value of acquiring information, which depends on the extent to which payoffs vary with the state of the world, and the marginal cost of acquiring information.

2.2 Optimal unconditional probabilities

The integral in (5) can be solved given the distributional assumptions that we have just introduced, and the constrained maximization problem simplifies to:[18]

$$\max_{\mathcal{P}_1,\ldots,\mathcal{P}_k} \ln\left[\sum_{a\in B_k} e^{\overline{v}_a+\lambda\ln(\mathcal{P}_a)}\right] \tag{7}$$

The maximization problem in (7) can be solved for an arbitrary set B_k, with $k = 1, ..., N$; the solution is given by:[19]

$$\mathcal{P}_h^{B_k} = \frac{e^{\overline{v}_h/(1-\lambda)}}{\sum_{a\in B_k} e^{\overline{v}_a/(1-\lambda)}} \tag{8}$$

We can show that the expected utility from choosing with positive probability the alternatives in the set B_k monotonically increases with k, so that the consideration set is given by $B_N = A$, i.e., all alternatives are always selected with positive probability.[20] The optimal unconditional probabilities are given by:

$$\mathcal{P}_h = \frac{e^{\overline{v}_h/(1-\lambda)}}{\sum_{a\in A} e^{\overline{v}_a/(1-\lambda)}} \tag{9}$$

[16]The variance σ^2 of $C(\lambda)$ is equal to $(1-\lambda^2)\pi^2/6$, so that the variance of $v_k + \lambda\epsilon_k$ is equal to $\pi^2/6$, i.e., the variance of a Gumbel distribution with a scale parameter equal to 1.

[17]As with a Gumbel distribution, the difference of two independent $C(\lambda)$ random variables follows a logistic distribution (Cardell, 1997), with scale parameter equal to $\sqrt{1-\lambda^2}$.

[18]See the proof in the Appendix A.1.

[19]See the proof in the Appendix A.2.

[20]See the proof in the Appendix A.3; this property also implies that, as in Brown and Jeon (2020), the model can still admit a closed-form solution in the presence of unobserved individual heterogeneity, as all alternatives are always included in the (individual) consideration set.

Notice, as proved by Dasgupta and Mondria (2018), that $\mathcal{P}_h$ is a non-monotonic function of λ for $h = 2, ..., N-1$, while $\mathcal{P}_1$ ($\mathcal{P}_N$) is monotonically increasing (decreasing) in λ, as:

$$\frac{\partial \ln \mathcal{P}_h}{\partial \lambda} = \frac{e^{\overline{v}_h/(\lambda-1)}}{(1-\lambda)^2}\left(\overline{v}_h - \sum_{a=1}^{N} \mathcal{P}_a \overline{v}_a\right) \tag{10}$$

The sign of the partial derivative in (10) depends on the difference between $\overline{v}_h$ and a probability-weighted average of the payoffs of all alternatives, which is unambiguously lower (higher) than the payoff of the most (least) attractive alternative. We will exploit the fact that:

$$\frac{\partial \ln \mathcal{P}_1}{\partial \lambda} > 0 \tag{11}$$

in the empirical analysis to obtain information on the unobserved value of this key parameter from observed past migration flow data.

2.3 Closed-form conditional choice probabilities

If we plug in (4) the expression for the unconditional choice probabilities in (9), when $B_k = A$:

$$\mathcal{P}_k(\boldsymbol{\epsilon}) = \frac{\mathcal{P}_k e^{v_k/\lambda}}{\sum_{a\in A} \mathcal{P}_a e^{v_a/\lambda}} = \frac{e^{\frac{\epsilon_k}{\lambda} + \frac{\overline{v}_k}{\lambda(1-\lambda)}}}{\sum_{a\in A} e^{\frac{\epsilon_a}{\lambda} + \frac{\overline{v}_a}{\lambda(1-\lambda)}}} \tag{12}$$

The conditional choice probability $\mathcal{P}_k(\boldsymbol{\epsilon})$ can be written as a function of the *un*conditional choice probabilities $\mathcal{P}_a$, which only depend on the vector of expected payoffs $\overline{\boldsymbol{v}}$, and on the vector $\boldsymbol{\epsilon}$ of the realizations of the deviation of the actual payoffs from their expected values.

2.4 Optimal cost of information acquisition

We can gain further insights on features of the solution of the location-choice problem with costly information acquisition by analyzing a simplified version of the model where $A = \{1, 2\}$, which gives us the opportunity to present the results graphically.[21] Without loss of generality, we can set $\overline{v}_2 = 0$; the optimal conditional probability $\mathcal{P}_1(x)$ of selecting

[21] We describe below how these results generalize to the case in which $N > 2$.

alternative 1 is thus given by:

$$\mathcal{P}_1(x) = \left[1 + \left(\frac{1-\mathcal{P}_1}{\mathcal{P}_1}\right)^{1/\lambda} e^{-x/\lambda}\right]^{-1} \tag{13}$$

where $x = \epsilon_1 - \epsilon_2$ follows a logistic distribution, with the cumulative distribution:

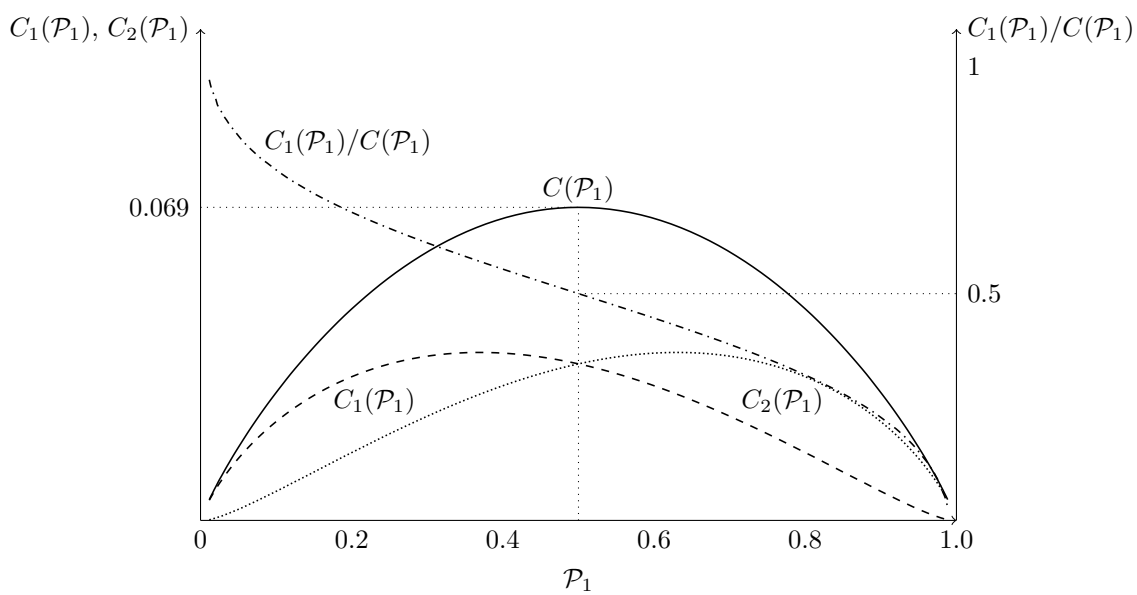
$$G(x) = \left(1 + e^{-x/\sqrt{1-\lambda^2}}\right)^{-1}.$$

We can thus rewrite the two alternative-specific costs of information acquisition as follows:

$$C_1(\mathcal{P}_1) = \lambda\left(-\mathcal{P}_1 \ln \mathcal{P}_1 + \int_{-\infty}^{+\infty} \mathcal{P}_1(x) \ln \mathcal{P}_1(x) g(x)\,\mathrm{d}x\right) \tag{14}$$

where $g(x) = \partial G(x)/\partial x$, and $C_2(\mathcal{P}_1) = C_1(1-\mathcal{P}_1)$.

Figure 1: Total and (absolute and relative) alternative-specific optimal cost of information



Notes: the integral that enters into the expression for $C_1(\mathcal{P}_1)$ is solved numerically for $\lambda = 0.1$.

The integrand function that appears in (14) does not admit a closed-form primitive, but we can gain insights on the total and alternative-specific investment in information

acquisition by numerically solving for $C_1(\mathcal{P}_1)$.[22] Figure 1 plots the values of $C(\mathcal{P}_1)$, $C_1(\mathcal{P}_1)$ and $C_2(\mathcal{P}_1)$ against $\mathcal{P}_1$ when $\lambda = 0.1$ (left-hand side vertical axis), as well as the value of the ratio $C_1(\mathcal{P}_1)/C(\mathcal{P}_1)$ (right-hand side axis).[23]

Several features of the evolution of the cost of information acquisition, and of its distribution between the two alternatives, with respect to $\mathcal{P}_1$, are worth emphasizing:[24] First, the total cost of information acquisition $C(\mathcal{P}_1)$ is maximized when $\mathcal{P}_1 = 1/2$,[25] and it is monotonically increasing (decreasing) in $\mathcal{P}_1$ when $\mathcal{P}_1 < 1/2$ ($\mathcal{P}_1 > 1/2$). Second, the migrant invests more in information acquisition about the alternative that is *a priori* less attractive, as $C_1(\mathcal{P}_1) < C_2(\mathcal{P}_1)$ when $\mathcal{P}_1 > \mathcal{P}_2$. Third, the alternative-specific investment in information acquisition is maximized when the probability of choosing an alternative is below 1/2. Fourth, the share of the total cost of information acquisition that is directed towards alternative 1, i.e., $C_1(\mathcal{P})/C(\mathcal{P})$, monotonically declines with the probability of selecting alternative 1. In terms of the signals, the migrant rationally decides to receive a more precise signal with respect to the payoff of the less attractive alternative, so that for this alternative the conditional choice probabilities vary more with respect to $x = \epsilon_1 - \epsilon_2$.

When the choice set A includes N alternatives, we can follow Bunch and Rocke (2016) to obtain independent draws of the payoffs from a $C(\lambda)$ distribution, and numerically compute the value of $C(\mathcal{P})$. This reveals that the properties that we have just described extend to an arbitrary number of alternatives. Notably, $C(\mathcal{P})$ is maximized when $\mathcal{P}_k = 1/N$, $\forall k \in A$,[26] and the cost of information acquisition for the alternative that is *a priori* most attractive is always below the cost of information acquisition for at least another alternative in the choice set A.[27]

[22] This is done by computing the value of the integral in (14) with 2,000 draws for x; the computation is repeated 2,000 times, and we then average $C_1(\mathcal{P}_1)$ over these replications; we then define $C_2(\mathcal{P}_1) = C_1(1-\mathcal{P}_1)$.

[23] We thank an anonymous referee for pushing us to explore the uneven allocation of attention across alternatives in the choice set.

[24] These properties are independent from the value of λ, and are demonstrated analytically when the two alternatives are *ex ante* identical in the Appendix A.4; an increase in λ exerts an ambiguous effect on $C(\mathcal{P}_1)$, while it unambiguously reduces the optimal reduction in the entropy of the payoffs, i.e., $C(\mathcal{P}_1)/\lambda$.

[25] We have that $C(1/2) = 0.069$; as $\lambda = 0.1$, the reduction in entropy stands at 0.69; as the entropy of the distribution of the priors is approximately equal to $2(1+\gamma) \approx 3.14$, where $1+\gamma$ is the entropy of a univariate Gumbel distribution, so the entropy is reduced by approximately 22 percent with the optimal signal acquisition strategy.

[26] We can also demonstrate that $C(\mathcal{P})$ monotonically increases with N, while $C_k(\mathcal{P}) = C(\mathcal{P})/N$, $\forall k \in A$, monotonically declines with the size of the choice set when alternatives are *ex ante* identical.

[27] A corollary of this property is that $C_k(\mathcal{P})$ is maximized when $\mathcal{P}_k < 1/2$, and we can demonstrate that $C_k(\mathcal{P})$ is an hump-shaped function of $\mathcal{P}_k$.

2.5 Elasticities

The semi-elasticity of the choice probability $\mathcal{P}_k(\boldsymbol{\epsilon})$ with respect ϵ_k and the expected value of this semi-elasticity are given by:

$$\frac{\partial \ln\left[\mathcal{P}_k(\boldsymbol{\epsilon})\right]}{\partial \epsilon_k} = \frac{1}{\lambda}\left[1-\mathcal{P}_k(\boldsymbol{\epsilon})\right], \; \mathbb{E}_{\boldsymbol{\epsilon}}\left(\frac{\partial \ln\left[\mathcal{P}_k(\boldsymbol{\epsilon})\right]}{\partial \epsilon_k}\right) = \frac{1}{\lambda}\left(1-\mathcal{P}_k\right) \tag{15}$$

Thus, this elasticity is higher for alternatives whose unconditional probability of being chosen is low; this can be related to how the alternative-specific investment in information acquisition $C_k(\mathcal{P})$ is related to the unconditional choice probability $\mathcal{P}_k$. We can write down the corresponding expressions for the elasticities with respect to $\overline{v}_k$:

$$\frac{\partial \ln\left[\mathcal{P}_k(\boldsymbol{\epsilon})\right]}{\partial \overline{v}_k} = \frac{1}{\lambda(1-\lambda)}\left[1-\mathcal{P}_k(\boldsymbol{\epsilon})\right], \; \mathbb{E}_{\boldsymbol{\epsilon}}\left(\frac{\partial \ln\left[\mathcal{P}_k(\boldsymbol{\epsilon})\right]}{\partial \overline{v}_k}\right) = \frac{1}{\lambda(1-\lambda)}\left(1-\mathcal{P}_k\right) \tag{16}$$

The ratio between (15) and (16) stands at $1-\lambda$: when λ increases, the relative size of the average elasticity of $\mathcal{P}_k(\boldsymbol{\epsilon})$ with respect to deviations of the payoff from its expected value declines, as the migrant is (rationally) receiving less precise signals about the payoff.

2.6 Testable implication

Our location-choice model with costly information acquisition implies that (i) the responsiveness of optimal conditional choice probabilities to variations in the expected value of the payoff in one alternative in the choice set is negatively related to λ, the parameter that determines the marginal cost of information acquisition, as shown in equation (15), and that (ii) there is a monotonic positive relationship between λ and $\mathcal{P}_1$, i.e., the unconditional probability of opting for the most attractive destination, as shown in (11). The econometric analysis will exploit point (ii) to build the empirical counterpart of λ from the distribution of past (origin-specific) international migration flows, and bring to the data the testable implication described at point (i).

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 the empirical proxy for the cost of information acquisition λ (or, more

precisely, for $1/\lambda$), which we will term 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 census-based data on the stock of individuals (by country of birth) residing in each country. More precisely, Abel (2018) recovers the minimal amount of bilateral flows that are required to match the observed evolution of stock data, once these have been 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, 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.[28] As discussed below, these two aspects are critical to generate from the data what we define as the value of information, the key variable that allows us to recover the effect of income at destination in a way that reflects the presence of information frictions in the location-decision problem that migrants face. 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.[29,30] The average value of m_{jkt} stands at 957.4, with a standard deviation of 15,472.4, and 61.2 percent of zero flows.

[28]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).

[29]Migration flows before 1980 are used to measure the value of information (see Section 3.2 below).

[30]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 minor destination countries, which represent less than 0.9 percent of total migration flows in Abel (2018).

3.2 Measurement of the value of information

Eq. (11) suggests that we can build from the data a suitable empirical counterpart of the (unknown) origin-specific value of information. The location-decision problem presented in Section 2 is static, while the availability of longitudinal data on bilateral migration flows allows us to build an empirical counterpart of the value of information that is possibly time-varying. More precisely, we proxy $\mathcal{P}_1$ with the share of migration flows directed from j to the main foreign destination in a period up to t. We rely on $p(r)_{jt}$, defined as follows:[31]

$$p(r)_{jt} \equiv \max_k \left\{ \frac{\sum_{t-r}^{t} m_{jkt}}{\sum_{t-r}^{t} \sum_{l \in A} m_{jlt}} \right\}, \; r = \{5, 10, 15, 20\} \tag{17}$$

It is interesting to note that 105 different countries represent the main destination, and hence determine the value of information, for at least one of the 1,347 origin-year pairs in our estimation sample; unsurprisingly, 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, thus revealing the importance of having a comprehensive set of destinations covered in the data. As $\mathcal{P}_1$ is a monotonically increasing function of λ, as demonstrated in (10), while the value of information is negatively related to λ, we measure it through the following transformation of $p(r)_{jt}$:[32]

$$w(r)_{jt} = -\ln[p(r)_{jt}] \tag{18}$$

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 implies that $w(5)_{\text{CHN1995}} = -\ln(0.254) = 1.371$. Thus,

[31] Notice that $p(r)_{jt}$ in (17) 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 $r = 20$.

[32] This specific functional form is immaterial for the evidence that we present in Sections 4.2 and 4.3, which is robust to interacting GDP per capita at destination with $p(r)_{jt}$, or with $1/p(r)_{jt}$; results are available from the Authors upon request.

the empirical counterpart of the value of information in (18) suggests that Chinese migrants valued information more than Mexican migrants in the five-year period starting in 1995.

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 (18).

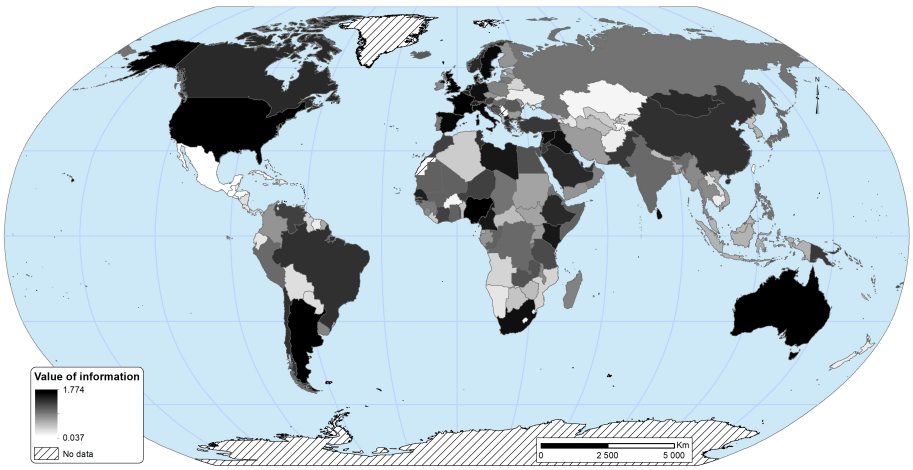
Source: Authors' elaboration on Abel (2018).

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. When we increase the length r of the time period over which we measure past migration flows, we get closer to the objective of obtaining a proxy for the *un*conditional probability of selecting the main alternative, but we also run the risk of introducing noise that is due to changes in the attractiveness of the various destinations; hence, it is important to test the robustness of our empirical evidence when cumulating past flows over different periods. Notice that, when r increases, the ensuing variation in $p(r)_{jt}$ can also reflect the change in the main destination: when we move from 5 to 10 years, we observe such a switch for 331 out of 1,347 origin-period pairs, and the corresponding figures for 15 and 20 years stand at 459 and 521 origin-period pairs. Nevertheless, the four variants of the empirical counterparts of the value of information are closely correlated: the correlations range between 0.58 (between $w(5)_{jt}$ and $w(20)_{jt}$) and 0.93 ($w(15)_{jt}$ and $w(20)_{jt}$). For $w(5)_{jt}$, the 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.[33] The variability in $w(5)_{jt}$ reflects both time-invariant differences across origins, as well as within-origin differences over time. More precisely, a

[33]The upper bound of the value of information stands at $-\ln 1/N = \ln 184 \approx 5.2$ when $N = 184$.

regression of $w(5)_{jt}$ on a set of origin dummies explains 40.4 percent of its variability. Beyond differences in λ, time-invariant heterogeneity across origins in $w(r)_{jt}$ might also capture the effects of geography, e.g., proximity to a high-income country increases the concentration of migration flows, while its within-origin variability might reflect as well variations in (observed or unobserved) determinants of the attractiveness or accessibility of major destinations.

Figure 2: Origin-specific average of the value of information $w(5)_{jt}$



Source: Authors' elaboration on Abel (2018).

Figure 2 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. Figure 2 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.

4 Econometric analysis

Our objective is to test the empirical relevance of information frictions in shaping migration decisions. To this end, we bring to the data a theory-based specification of the migration gravity equation where we introduce an interaction between the empirical counterpart of the

value of information and income per capita at destination.

4.1 Gravity equation with rational inattention

We can write the migration flows m_{jkt} between an origin j and a destination k in the five-year period starting in year t as:

$$m_{jkt} = \mathcal{P}_{jkt} \times n_{jt} \times \zeta_{jkt} \tag{19}$$

where $n_{jt} = \sum_{k \in A} m_{jkt}$, $\zeta_{jkt} > 0$ is an error term, and the probability $\mathcal{P}_{jkt}$ that destination k represents the utility-maximizing alternative for a migrant from j in period t is given by (12). Replacing $\mathcal{P}_{jkt}$, we can then rewrite equation (19) as:

$$m_{jkt} = \exp\left[\frac{1}{\lambda_{jt}}\epsilon_{kt} + \frac{1}{\lambda_{jt}(1-\lambda_{jt})}\overline{v}_{kt} + \Omega_{jt} + \ln(\zeta_{jkt})\right] \tag{20}$$

where:

$$\Omega_{jt} \equiv \ln(n_{jt}) - \sum_{a \in A} e^{\frac{\epsilon_{jat}}{\lambda_{jt}} + \frac{\overline{v}_{jat}}{\lambda_{jt}(1-\lambda_{jt})}}$$

We assume that the destination-specific utility $v_{jkt} = \overline{v}_{jkt} + \epsilon_{jkt}$ follows:

$$v_{jkt} = \alpha \ln\left(\frac{y_{kt}}{\tau_{jkt}}\right) \tag{21}$$

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 (21) implies 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 further assume that $\overline{v}_{jkt} = -\alpha \ln \tau_{jkt}$ and $\epsilon_{jkt} = \alpha \ln y_{kt}$, i.e., migrants can observe the determinants of the accessibility of destination k, but are unable to costlessly observe local economic conditions. These assumptions allow rewriting (20) as follows:

$$m_{jkt} = \exp\left[\frac{\alpha}{\lambda_{jt}} \ln y_{kt} - \frac{\alpha}{\lambda_{jt}(1-\lambda_{jt})} \ln \tau_{jkt} + \Omega_{jt} + \ln(\zeta_{jkt})\right] \tag{22}$$

The specification that we bring to the data is given by:

$$m_{jkt} = \exp\left[\beta\left(w(r)_{jt} \times \ln y_{kt}\right) + d_{kt} + d_{jt} + d_{jk} + \varepsilon_{jkt}\right] \tag{23}$$

where $w(r)_{jt}$, with $r = \{5, 10, 15, 20\}$, represents, as discussed in Section 3.2, an empirical proxy for $1/\lambda_{jt}$; $\ln y_{kt}$ is the logarithm of GDP per capita in 2010 USD from World Bank (2018);[34] d_{kt}, d_{jt}, and d_{jk} represent destination-time, origin-time and origin-destination (dyadic) dummies; and ε_{jkt} is the error term. Since we have a large share of zeros (61.2 percent) in our dependent variable m_{jkt}, we estimate (23) 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. (2019, 2020), which allows handling in a computationally efficient way the large number of fixed effects in (23). Standard errors are clustered at the origin level following Bertrand et al. (2004).

The inclusion of origin-time dummies in (23) perfectly controls for Ω_{jt} in (20). The rich structure of fixed effects allows controlling for the dependence of the iceberg-type migration costs τ_{jkt} on dyadic time-invariant factors such as geographical distance, linguistic and cultural proximity, or on destination-time specific factors, such as policy-induced barriers to migrations. However, (22) reveals that the effect of $\ln \tau_{jkt}$ on bilateral migration flows is also mediated by λ_{jt}, so this confounding effect is potentially specific to each origin-destination-time triplet. We pursue two different and not mutually exclusive approaches to control for it: first, we augment the specification in (23) by interacting typical correlates of dyadic migration costs from Mayer and Zignago (2011) with $w(r)_{jt}$; second, we also control for $\ln(s_{jkt} + 1)$, i.e., the logarithm of (one plus) the stock of j-born migrants residing in destination k in year t, as in Beine et al. (2011).[35,36]

Our estimate for β will be consistent as long as (i) the effect of immigration flows from one particular origin on one particular destination is close to zero, and (ii) our proxy for information costs is both predetermined and persistent enough. Under (i) and (ii), there will be no simultaneity between our dependent variable, migration flows, GDP per capita and our empirical value of information. Condition (i) is likely to be satisfied. For example, the median migration flow in our dataset amounts to 0.003 per cent of the destination country population in a particular five-year period. According to the findings in Ortega and Peri

[34] The average and standard deviation of $\ln y_{kt}$ over our sample stand at 8.24 and 1.53 respectively.

[35] 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 stand at 2.25 and 2.95 respectively.

[36] The econometric evidence is fully robust when relying on the inverse hyperbolic sine transformation of s_{jkt} to account for zeros in bilateral migrant stocks, or when also interacting the measure of networks with $w(r)_{jt}$; results are available from the Authors upon request.

(2014), this would translate into an increase in the GDP per capita of the typical destination country of 0.02 per cent over five years, that is, barely 0.004 per cent per year. As far as (ii) is concerned, the condition is clearly satisfied in theory, as λ is a parameter that determines migration flows. When it comes to our empirical proxy, $w(r)_{jt}$ is calculated on past migration flows and we experiment with different values of r precisely to make sure that our results hold under different notions of persistence. Furthermore, it must be emphasized that the empirical value of information $w(r)_{jt}$ is not a lagged version of the dependent variable. Recall that our dependent variable includes variation at the origin-destination-time level (jkt) while the value of information is origin-time specific (jt). All past flows out of an origin enter into the computation of $w(r)_{jt}$ but most of its variation corresponds to the main destination out of 189 in our dataset. This is why we show below that our results are robust to dropping the main destination, for which the lagged dependent variable would create a problem. Still, any remaining auto-correlation should be taken into account by our clustering of standard errors at the origin level.

4.2 Main results

Table 2 reports our baseline results for the gravity equation described in (23). Each 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. 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 conventional confidence levels.[37] 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 0.072, in column (1), and 0.093, in column (4). Going back to the example of China and Mexico that we introduced in Section 3.2, the estimates in Table 2 imply that the elasticity for migration from China to any destination between 1995 and 2000 was 0.182-0.241 higher than the corresponding elasticity for migration from Mexico over the same time period. Similarly, the estimates also imply a substantial variability over time for a given origin; for instance, the elasticity of migration out of Ecuador increased by 0.078-0.104 between the early 1980s

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

and the early 2000s,[38] following a substantial diversification of the main destinations for Ecuadorian migrants (Bertoli et al., 2011).

Table 2: Baseline results

	Dependent variable: m_{jkt}			
	(1)	(2)	(3)	(4)
Value of r	5	10	15	20
$\ln(y_{kt}) \times w(r)_{jt}$	0.136**	0.166**	0.163**	0.180*
	(0.058)	(0.066)	(0.080)	(0.101)
Observations	221,342	224,184	225,327	225,458
Pseudo-R^2	0.962	0.962	0.962	0.962
$w(r)_{jt}$ (mean)	0.863	0.922	0.950	0.965
$w(r)_{jt}$ (s.d.)	0.533	0.524	0.519	0.517
d_{jt}, d_{kt} and d_{jk}	Yes	Yes	Yes	Yes

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$; clustered standard errors at the origin level in parentheses.
Source: Authors' elaboration on Abel (2018) and World Bank (2018).

We next show how our main results are affected when we perform some particular variations in the estimated specifications, and how they consolidate the interpretation that the value of information picks up information frictions in the way that our simple model describes.

4.2.1 Dropping the main destination from the estimation sample

First, we address the concern described at the end of Section 4.1. Since the value of information is constructed using lags of the dependent variable for all destinations, we check whether our results are robust to dropping the main origin-time specific destination from the estimation sample, as past flows to this specific country pick up, by construction, most of the variation in the value of information. The exercise is performed in Table 3. We can see that our estimate for β decreases in size for all four definitions of the value of information. Still, the coefficients remain significant at conventional levels (and more precisely estimated), and statistically identical to our main results in Table 2.

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

Table 3: Results excluding the main destination from the sample

	Dependent variable: m_{jkt}			
	(1)	(2)	(3)	(4)
Value of r	5	10	15	20
$\ln(y_{kt}) \times w(r)_{jt}$	0.094***	0.117***	0.122***	0.108*
	(0.033)	(0.034)	(0.042)	(0.055)
Observations	220,088	222,912	224,046	224,180
Pseudo-R^2	0.954	0.954	0.955	0.956
$w(r)_{jt}$ (mean)	0.863	0.922	0.950	0.965
$w(r)_{jt}$ (s.d.)	0.533	0.524	0.519	0.517
d_{jt}, d_{kt} and d_{jk}	Yes	Yes	Yes	Yes

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$; clustered standard errors at the origin level in parentheses.
Source: Authors' elaboration on Abel (2018) and World Bank (2018).

Besides showing that the results are not mechanically generated by lagged migration flows, the estimates in Table 3 prove that ours is not a story about Mexican migration flows to the United States being less responsive to economic conditions in the United States. Mexican migration flows are less responsive to economic conditions also in other destinations than, for example, Chinese emigration flows.

4.2.2 Using the past share of flows to the the second destination

The analysis of the theoretical model has revealed that only the unconditional probability of opting for the main destination is monotonically related to $1/\lambda_{jt}$, while the relationship of this key parameter of the model with the unconditional probabilities for other destinations is ambiguous. Thus, we define an alternative measure $w_2(r)_{jt} \equiv -\ln[p_2(r)_{jt}]$, where $p_2(r)_{jt}$ is the share of migrants to the second main destination rather than the share of migrants to the top destination in the past r years. We interact this alternative measure with the log of GDP per capita at destination in Table 4. This change in our variable of interest renders all of our estimated coefficients statistically insignificant. These results (or, rather, this lack of results) is fully consistent with our theoretical model.[39]

[39]Similar evidence is obtained when using the past share of flows directed to the third, fourth or fifth destination; results are available from the Authors upon request.

Table 4: Measuring the value of information with the second main destination

	Dependent variable: m_{jkt}			
	(1)	(2)	(3)	(4)
Value of r	5	10	15	20
$\ln(y_{kt}) \times w_2(r)_{jt}$	-0.019	-0.109	-0.061	-0.045
	(0.030)	(0.067)	(0.084)	(0.085)
Observations	219,079	223,547	224,669	225,144
Pseudo-R^2	0.963	0.962	0.962	0.962
$w(r)_{jt}$ (mean)	2.061	2.034	2.035	2.032
$w(r)_{jt}$ (s.d.)	0.848	0.689	0.625	0.598
d_{jt}, d_{kt} and d_{jk}	Yes	Yes	Yes	Yes

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$; clustered standard errors at the origin level in parentheses. $w_2(r)_{jt}$ is equal to minus the logarithm of the share of past flows directed towards the second main destination.

Source: Authors' elaboration on Abel (2018) and World Bank (2018).

4.2.3 Controlling more thoroughly for migration costs

We advanced above, when discussing equation (22), two different strategies to control for the influence of migration costs τ_{jkt} on migration flows, as their effect was also mediated by information costs λ_{jt}. First, Table 5 expands our preferred specification by controlling for the interaction between the value of information and the classic dyadic time-invariant gravity determinants of migration flows: contiguity between j and k, the existence of a common language between j and k, whether j and k ever had a common colonial link, and the logarithm of the geodesic distance between j and k. Most of the interactions of these added variables with the value of information turn out not to be significant. We have one significant positive interaction of distance out of four and two marginally significant negative interactions of the colony variable. In contrast, our interaction of interest between the value of information and GDP per capita at destination remains positive and significant and, while all the coefficients go down in size with respect to our baseline in Table 2, the differences between both sets of coefficients are not statistically significant.

Table 5: Interactions of dyadic variables with $w(r)_{jt}$

	Dependent variable: m_{jkt}			
	(1)	(2)	(3)	(4)
Value of r	5	10	15	20
$\ln(y_{kt}) \times w(r)_{jt}$	0.092**	0.131***	0.119*	0.167**
	(0.045)	(0.050)	(0.063)	(0.081)
$w(r)_{jt} \times$ Contiguity$_{jk}$	-0.107	-0.212	-0.238	0.025
	(0.153)	(0.154)	(0.189)	(0.279)
$w(r)_{jt} \times$ Common language$_{jk}$	0.098	-0.093	0.035	0.033
	(0.091)	(0.101)	(0.103)	(0.146)
$w(r)_{jt} \times$ Colony$_{jk}$	-0.192*	-0.157	-0.256*	-0.230
	(0.111)	(0.134)	(0.141)	(0.162)
$w(r)_{jt} \times \ln(\text{distance}_{jk})$	0.109***	0.033	0.061	0.059
	(0.035)	(0.041)	(0.050)	(0.073)
Observations	214,838	217,518	218,654	218,785
Pseudo-R^2	0.963	0.962	0.962	0.962
$w(r)_{jt}$ (mean)	0.867	0.928	0.957	0.972
$w(r)_{jt}$ (s.d.)	0.531	0.524	0.520	0.517
d_{jt}, d_{kt} and d_{jk}	Yes	Yes	Yes	Yes

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$; clustered standard errors at the origin level in parentheses.

Source: Authors' elaboration on Abel (2018), World Bank (2018) and Mayer and Zignago (2011).

Second, in Table 6, we augment our baseline specification with the variable $\ln(s_{jkt} + 1)$, the logarithm of (one plus) the stock of j-born migrants residing in destination k in year t, as in Beine et al. (2011). This serves two purposes. On the one hand, it allows us to control directly for an observable factor that has been shown to be relevant in affecting migration costs (McKenzie and Rapoport, 2010). On the other hand, it shows that our value of information is not picking up omitted network effects. While the coefficient for the stock of previous migrants from the same destination is positive and highly significant in all specifications, our estimated $\widehat{\beta}$ also remains positive and significant, and close to our baseline estimates in Table 2.

Table 6: Baseline results on the value of information with networks

	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*
	(0.051)	(0.057)	(0.070)	(0.089)
$\ln(s_{jkt} + 1)$	0.193***	0.192***	0.195***	0.196***
	(0.032)	(0.031)	(0.031)	(0.031)
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: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$; clustered standard errors at the origin level in parentheses.
Source: Authors' elaboration on Abel (2018), World Bank (2018), Özden et al. (2011) and United Nations Population Division (2015a).

In Table 7, we also interact the network variable with our variable for the value of information, as equation (22) suggests that any component of migration costs will have its effect on migration flows mediated through information costs. Table 7 shows that the interaction between the network variable and the value of information is not significantly different from zero. This is not surprising considering that the elasticity of migration flows with respect to migration costs, while depending on λ_{jt}, is not monotonic in λ_{jt}. On the contrary, the elasticity of migration flows with respect to GDP per capita at destination is monotonically related to λ_{jt} and this is reflected in the positive and statistically significant coefficient $\widehat{\beta}$ in all specifications in Table 7. Again, these coefficients are not statistically different from those reported in the baseline. Our results are fully robust when we put together both strategies for more thoroughly controlling for migration costs, that is, when we combine Tables 5 and 7 and include both the network variable, its interaction with the value of information and the interactions of the value of information with time-invariant dyadic variables.[40]

[40]Results are available from the Authors upon request.

Table 7: Interacting networks with the value of information

	Dependent variable: m_{jkt}			
	(1)	(2)	(3)	(4)
Value of r	5	10	15	20
$\ln(y_{kt}) \times w(r)_{jt}$	0.122**	0.146**	0.144**	0.157*
	(0.051)	(0.058)	(0.072)	(0.091)
$\ln(s_{jkt} + 1) \times w(r)_{jt}$	-0.017	-0.011	-0.001	0.023
	(0.017)	(0.021)	(0.024)	(0.028)
$\ln(s_{jkt} + 1)$	0.207***	0.201***	0.196***	0.175***
	(0.038)	(0.042)	(0.044)	(0.047)
Observations	220,627	223,469	224,612	224,743
Pseudo-R^2	0.964	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: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$; clustered standard errors at the origin level in parentheses.
Source: Authors' elaboration on Abel (2018), World Bank (2018), Özden et al. (2011) and United Nations Population Division (2015a).

4.2.4 Do migrants form consideration sets?

Our theoretical model implies that all choice probabilities should be strictly positive. Under the assumptions that we needed to invoke to provide an analytical solution for our model there would be no zero flows. However, empirically we observe that 61.2 per cent of observations correspond to zero migration flows over a five-year period. Discrete-choice models of rational inattention can lead, under alternative distributional assumptions, to the formation of consideration sets that are strictly smaller than the choice set (Caplin et al., 2019).

In this spirit, 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 60.4 percent of the observations in our sample correspond to origin-destination dyads with a zero flow in the recent past. Notice that we do not even use 37 per cent of these for identification since they correspond to origin-destination pairs where the flows are always zero in our baseline sample. Still, we would not want our result on the value of information, derived from a model where zero flows are not possible, to be affected by these zero-flow observations. Intuitively, the migration flows for these dyads could be

less sensitive to variations in economic conditions in the various destination countries, as migrants from j could exclude destination k from their (time-varying) consideration sets when $d^{\text{zero}}(5)_{jkt} = 1$.

Table 8: Zero past flows reduce current responsiveness

	Dependent variable: m_{jkt}		
	(1)	(2)	(3)
Value of r	5	5	5
$\ln(y_{kt}) \times d^{\text{zero}}(r)_{jkt}$	-0.039***	-0.037***	
	(0.007)	(0.007)	
$\ln(y_{kt}) \times w(r)_{jt}$		0.132**	0.136**
		(0.059)	(0.058)
Observations	221,342	221,342	221,342
Pseudo-R^2	0.962	0.963	0.962
$w(r)_{jt}$ (mean)	0.863	0.863	0.863
$w(r)_{jt}$ (s.d.)	0.533	0.533	0.533
d_{jt}, d_{kt} and d_{jk}	Yes	Yes	Yes

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$; clustered standard errors at the origin level in parentheses.
Source: Authors' elaboration on Abel (2018) and World Bank (2018).

Table 8 confirms that this is indeed the case: the elasticity with respect to GDP per capita at destination is 0.039 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 columns in Table 8 reveals. This also applies when using data over the previous 10, 15 or 20 years to identify origin-destination pairs with past zero flows, or when we define a relative or an absolute threshold higher than zero to identify minor destinations.[41]

[41] Results are available from the Authors upon request.

4.3 Threats to our interpretation

The econometric evidence presented in Section 4.2 above is consistent with the testable implications laid out in Section 2.6, but we need to understand whether they could also be generated by a canonical full-information model, or by a full-information model with a richer and more flexible specification of location-specific utility.

What would it happen if migrants were able to costlessly observe location-specific utilities before deciding where to move? A random utility maximization model with distributional assumptions *à la* McFadden, and where the variance of the stochastic component of utility is origin-specific, also implies a systematic relationship between the distribution of migrants across destinations and the responsiveness of bilateral migration flows with respect to variations in economic conditions of the various destinations.[42] More precisely, origin countries with a greater preference heterogeneity will have migration flows that are both (i) more dispersed across destinations, and (ii) less responsive to changes in economic conditions. This, in turn, implies that a canonical full information model generates the testable implication that the coefficient of the interaction between $w(r)_{jt}$ and $\ln y_{kt}$ should be negative, a prediction that is clearly rejected by the data.

The pattern that we uncover in the data might be explained by a more flexible version of the full-information model. For instance, migration decisions could be subject to binding liquidity constraints, which could influence migrants' ability to respond to variations in economic conditions 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} (an assumption that we have retained so far), 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.[43]

4.3.1 Liquidity constraints

The empirical counterpart $w(5)_{jt}$ for the value of information is higher in some geographical areas where most high-income countries are concentrated (see Figure 2 in Section 3.2). If we

[42] The full analysis of this model is presented in the Appendix B.1.

[43] Porcher (2019) provides empirical evidence that bilateral migration flows respond more to economic conditions in the various destinations for closer origin-destination pairs.

rely on the classification by income groups from the World Bank, we have that the average value of $w(5)_{jt}$ is equal to 1.088 for high-income origin countries, and to 0.805 for the other origin countries.[44] Migration decisions can be subject to binding liquidity constraints, as shown notably by Clemens (2014), Angelucci (2015), Djajic et al. (2016), Bazzi (2017) or Dao et al. (2018).

Table 9: Heterogeneity by income group

	Dependent variable: m_{jkt}			
	(1)	(2)	(3)	(4)
Value of r	5	10	15	20
$\ln(y_{kt}) \times w(r)_{jt}$	0.136**	0.162**	0.159*	0.180*
	(0.059)	(0.067)	(0.081)	(0.104)
$\ln(y_{kt}) \times d_{jt}^{\text{low}}$	-0.163	-0.136	-0.137	-0.124
	(0.124)	(0.124)	(0.128)	(0.133)
$\ln(y_{kt}) \times d_{jt}^{\text{l. middle}}$	-0.206*	-0.180*	-0.194*	-0.189*
	(0.110)	(0.109)	(0.112)	(0.115)
$\ln(y_{kt}) \times d_{jt}^{\text{u. middle}}$	-0.051	-0.039	-0.057	-0.047
	(0.085)	(0.085)	(0.086)	(0.090)
Observations	216,742	219,584	220,727	220,858
Pseudo-R^2	0.962	0.962	0.962	0.962
$w(r)_{jt}$ (mean)	0.864	0.929	0.962	0.978
$w(r)_{jt}$ (s.d.)	0.528	0.521	0.516	0.513
d_{jt}, d_{kt} and d_{jk}	Yes	Yes	Yes	Yes

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$; clustered standard errors at the origin level in parentheses.
Source: Authors' elaboration on Abel (2018) and World Bank (2018).

Liquidity constraints imply that the set of affordable destinations is smaller than the choice set (Marchal and Naiditch, 2020), and hence this pattern in the data poses a threat to our interpretation of the results in Table 2. Migrants from lower-income countries might not value information less, but they might be less able to react to variations in economic conditions, and their past distribution could be more concentrated in the main (affordable)

[44]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).

destination. We thus estimate an extended version of the gravity equation in (23), where we allow for a heterogeneous effect of $\ln(y_{kt})$ across groups of origins with a different level of income. Table 9 reveals that the elasticity of the migration rate with respect to y_{kt} is higher for origins classified as high-income countries in year t (the omitted category), albeit these differences are not precisely identified.[45] However, this does not influence either the size or the significance of the coefficient for our interaction effect, thus dismissing the concern that the values of $\widehat{\beta}$ in Table 2 were picking up a spurious correlation between $w(r)_{jt}$ and the income group to which the origin j belonged in year t.[46]

4.3.2 More flexible responsiveness to economic conditions

Do the results presented in Table 2 survive once we allow for a more general functional form of the deterministic component of utility v_{jkt}, or for differences across destinations or at the dyadic level in the cost of acquiring information, thus relaxing the assumption that λ does not vary across alternatives in the choice set? 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.[47] 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 four data columns in Table 10, in line with the idea that larger past migration flows reduce the cost of acquiring information on the attractiveness of the alternative destinations.[48] However, the inclusion of the additional 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 0.136 and 0.180 in Table 2, and between 0.112 and 0.193 in Table 10.

[45] Notice that liquidity constraints can hinder the ability of migrants to react to an increase in the attractiveness of a country, but they do not limit their ability to react to worsening economic conditions.

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

[47] The empirical counterparts for λ are insensitive to the scale of past migration flows.

[48] An alternative, but not mutually exclusive, explanation is that migrants' remittances help relaxing liquidity constraints at origin, thus increasing, as suggested by Table 10, the responsiveness of bilateral migration flows with respect to varying economic conditions.

Table 10: Interaction with the past emigration rate

	Dependent variable: m_{jkt}			
	(1)	(2)	(3)	(4)
Value of r	5	10	15	20
$\ln(y_{kt}) \times w(r)_{jt}$	0.112**	0.158**	0.151**	0.193**
	(0.050)	(0.062)	(0.073)	(0.097)
$\ln(y_{kt}) \times \ln[\text{emigration rate}(r)_{jt}]$	0.021**	0.017	0.048**	0.077***
	(0.009)	(0.024)	(0.020)	(0.024)
Observations	214,260	216,390	216,805	216,570
Pseudo-R^2	0.963	0.963	0.963	0.963
$w(r)_{jt}$ (mean)	0.866	0.931	0.966	0.982
$w(r)_{jt}$ (s.d.)	0.528	0.520	0.516	0.513
d_{jt}, d_{kt} and d_{jk}	Yes	Yes	Yes	Yes

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$; clustered standard errors at the origin level in parentheses.

Source: Authors' elaboration on Abel (2018) and World Bank (2018).

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 11. 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 those 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 12 similarly extends the gravity equation in (23) by introducing (either separately or jointly) interactions between the canonical dyadic controls from Mayer and Zignago (2011) and $\ln y_{kt}$: origin-destination pairs with lower dyadic migration costs, e.g., contiguous countries, are characterized by a greater responsiveness of bilateral migration flows with respect to economic conditions.[49] In particular, the interaction between $\ln y_{kt}$ and the geodesic distance between the origin j and the destination k is negative and significant. However, this does not influence the estimated coefficient for $\ln(y_{kt}) \times w(r)_{jt}$, which ranges between 0.141

[49]This result could be of independent interest with respect to the reliance on the estimation of a zero-stage gravity equation with dyadic time-invariant correlates of migration costs to generate an instrument for observed immigration (see, for instance, Ortega and Peri, 2014 and Alesina et al., 2016).

and 0.181, perfectly in line with the 0.136-0.180 range for $\widehat{\beta}$ from Table 2.[50]

Table 11: 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*
	(0.042)	(0.047)	(0.060)	(0.075)
$\ln(y_{kt}) \times \ln(s_{jkt}+1)$	-0.070***	-0.066***	-0.066***	-0.066***
	(0.018)	(0.018)	(0.018)	(0.017)
$\ln(s_{jkt}+1)$	0.845***	0.802***	0.808***	0.807***
	(0.186)	(0.181)	(0.184)	(0.180)
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: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$; clustered standard errors at the origin level in parentheses.
Source: Authors' elaboration on Abel (2018), World Bank (2018), 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, or bilateral, e.g., networks, correlates of the educational composition of migration flows (see, for instance, Beine et al., 2011) allows, at least partially, to downplay this concern.

[50] Additional results, which are available from the Authors upon request, reveal that our empirical evidence is also robust to interacting $\ln y_{kt}$ with various measures of cultural and linguistic proximity between the origin j and the destination k from Spolaore and Wacziarg (2016) and Adserà and Pytliková (2015).

Table 12: Heterogeneity with respect to dyadic determinants of migration costs

	Dependent variable: m_{jkt}			
	(1)	(2)	(3)	(4)
Value of r	5	10	15	20
$\ln(y_{kt}) \times w(r)_{jt}$	0.141**	0.173***	0.170**	0.183*
	(0.058)	(0.065)	(0.078)	(0.100)
$\ln(y_{kt}) \times$ Contiguity$_{jk}$	0.276	0.261	0.220	0.208
	(0.186)	(0.186)	(0.183)	(0.184)
$\ln(y_{kt}) \times$ Common language$_{jk}$	0.163	0.171	0.198	0.195
	(0.211)	(0.206)	(0.205)	(0.206)
$\ln(y_{kt}) \times$ Colony$_{jk}$	-0.347	-0.354	-0.371	-0.357
	(0.296)	(0.295)	(0.296)	(0.299)
$\ln(y_{kt}) \times \ln(\text{distance}_{jk})$	-0.203**	-0.213**	-0.225***	-0.222**
	(0.084)	(0.085)	(0.086)	(0.088)
Observations	214,838	217,518	218,654	218,785
Pseudo-R^2	0.963	0.963	0.962	0.962
$w(r)_{jt}$ (mean)	0.867	0.928	0.957	0.972
$w(r)_{jt}$ (s.d.)	0.531	0.524	0.520	0.517
d_{jt}, d_{kt} and d_{jk}	Yes	Yes	Yes	Yes

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$; clustered standard errors at the origin level in parentheses.

Source: Authors' elaboration on Abel (2018), World Bank (2018) and Mayer and Zignago (2011).

5 Concluding remarks

The insights obtained from applying the theory of rational inattention to the location-decision problem that migrants face are relevant to enhance our understanding of the determinants of international migration flows. The model delivers clear testable implications with respect to the role played by economic conditions in the various destinations in shaping incoming flows from origins that differ with respect to the value that migrants (rationally) attach to information acquisition. The theory reveals that the past distribution of origin-specific migration flows across destinations is informative about the (unknown) value of information. The econometric evidence is consistent with this testable prediction, and robust to alternative explanations derived from a model without information frictions.

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.

Alesina, A., J. Harnoss, and H. Rapoport (2016): "Birthplace diversity and economic prosperity," *Journal of Economic Growth*, 21, 101–138.

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

Bartoš, V., M. Bauer, J. Chytilová, and F. Matějka (2016): "Attention Discrimination: Theory and Field Experiments with Monitoring Information Acquisition," *American Economic Review*, 106, 1437–1475.

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.

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.

Bertrand, M., E. Duflo, and S. Mullainathan (2004): "How much should we trust differences-in-differences estimates?" *Quarterly Journal of Economics*, 119, 249–275.

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

Brown, Z. Y. and J. Jeon (2020): "Endogenous Information and Simplifying Insurance Choice," Mimeo, University of Michigan.

Bunch, D. S. and D. M. Rocke (2016): "Variance-component-based nested logit specifications: Improved formulation, and practical microsimulation of random disturbance terms," *Journal of Choice Modelling*, 21, 30–35.

Caplin, A., M. Dean, and J. Leahy (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 (2019): "The Place Premium: Bounding the Price Equivalent of Migration Barriers," *Review of Economics and Statistics*, 101, 201–213.

Correia, S., P. Guimarães, and T. Zylkin (2020): "ppmlhdfe: Fast Poisson Estimation with High-Dimensional Fixed Effects," *Stata Journal*, 20, 95–115.

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

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.

Djajic, S., M. G. Kirdar, and A. Vinogradova (2016): "Source-country earnings and emigration," *Journal of International Economics*, 99, 46–67.

Domencich, T. and D. L. McFadden (1975): *Urban Travel Demand: A Behavioral Analysis*, Oxford: North-Holland Publishing.

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

Fosgerau, M., E. Melo, A. de Palma, and M. Shum (2020): "Discrete Choice and Rational Inattention: a General Equivalence Result," *International Economic Review*, https://doi.org/10.1111/iere.12469.

Giordani, P. E. and M. Ruta (2013): "Coordination failures in immigration policy," *Journal of International Economics*, 89, 55–67.

Grogger, J. and G. H. Hanson (2011): "Income maximization and the selection and sorting of international migrants," *Journal of Development Economics*, 95, 42–57.

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 (2020): "How Borrowing Constraints Hinder Migration: Theoretical Insights from a RUM Model," *Scandinavian Journal of Economics*, 122, 732–761.

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.

Mayda, A. M. (2010): "International migration: A panel data analysis of the determinants of bilateral flows," *Journal of Population Economics*, 23, 1249–1274.

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. and H. Rapoport (2010): "Self-Selection Patterns in Mexico-U.S. Migration: The Role of Migration Networks," *The Review of Economics and Statistics*, 92, 811–821.

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.

Munshi, K. (2003): "Networks in the modern economy: Mexican migrants in the US labor market," *The Quarterly Journal of Economics*, 118, 549–599.

Natenzon, P. (2019): "Random Choice and Learning," *Journal of Political Economy*, 127, 419–457.

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

——— (2014): "Openness and income: The roles of trade and migration," *Journal of International Economics*, 92, 231–251.

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.

Porcher, C. (2019): "Migration with Costly Information," Mimeo, Princeton University.

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.

Shrestha, M. (2020): "Get Rich or Die Tryin': Perceived Earnings, Perceived Mortality Rates, and Migration Decisions of Potential Work Migrants from Nepal," *The World Bank Economic Review*, 34, 1–27.

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.

Sjaastad, L. A. (1962): "The Costs and Returns of Human Migration," *Journal of Political Economy*, 70, 80–93.

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.

Rational inattention and migration decisions

Simone Bertoli, Jesús Fernández-Huertas Moraga and Lucas Guichard

Appendix

A Proofs

A.1 Simplifying the maximization problem

The objective function in the constrained maximization problem that identifies the optimal choice probabilities within the set B_k is given by:

$$\int_{\boldsymbol{v}} \lambda \ln \left[\sum_{a \in B_k} \mathcal{P}_a e^{v_a/\lambda} \right] f(\boldsymbol{v}) \, \mathrm{d}\boldsymbol{v} \tag{A.1}$$

The key of the proof, which draws on Brown and Jeon (2020), rests on a result established by Domencich and McFadden (1975): in RUM models with full information and where the stochastic component of utility is i.i.d. EVT-1, we have that the expected value from the choice situation is equal to the logarithm of the sum of the exponentials of the expected value of utility in each alternative. Rewrite the objective function:

$$\begin{aligned}
\int_{\boldsymbol{v}} \lambda \ln \left[\sum_{a \in B_k} \mathcal{P}_a e^{v_a/\lambda} \right] f(\boldsymbol{v}) \, \mathrm{d}\boldsymbol{v} &= \int_{\boldsymbol{v}} \lambda \ln \left[\sum_{a \in B_k} e^{v_a/\lambda + \ln(\mathcal{P}_a)} \right] f(\boldsymbol{v}) \, \mathrm{d}\boldsymbol{v} \\
&= \lambda \mathbb{E}_{\boldsymbol{v}} \left[\ln \left(\sum_{a \in B_k} e^{v_a/\lambda + \ln(\mathcal{P}_a)} \right) \right] \\
&= \lambda \mathbb{E}_{\boldsymbol{\epsilon}} \left[\ln \left(\sum_{a \in B_k} e^{\overline{v}_a/\lambda + \ln(\mathcal{P}_a) + \epsilon_a/\lambda} \right) \right] \\
&= \lambda \mathbb{E}_{\boldsymbol{\epsilon},\boldsymbol{\eta}} \left[\max_{a \in B_k} \left(\overline{v}_a/\lambda + \ln(\mathcal{P}_a) + \epsilon_a/\lambda + \eta_a \right) \right]
\end{aligned}$$

where η_a is i.i.d EVT-1. If ϵ_a follows a $C(\lambda)$ distribution, then $\epsilon'_a \equiv \epsilon_a + \lambda\eta_a$ follows an EVT-1 distribution with scale parameter equal to 1. This implies that:

$$\int_{\boldsymbol{v}} \lambda \ln\left[\sum_{a\in B_k} \mathcal{P}_a e^{v_a/\lambda}\right] f(\boldsymbol{v})\,\mathrm{d}\boldsymbol{v} = \lambda \mathbb{E}_{\boldsymbol{\epsilon'}}\left[\max_{a\in B_k}\left(\overline{v}_a + \lambda\ln(\mathcal{P}_a) + \epsilon'_a\right)\right]$$
$$= \ln\left[\sum_{a\in B_k} e^{\overline{v}_a + \lambda\ln(\mathcal{P}_a)}\right]$$

A.2 Solving for optimal unconditional probabilities

The maximization problem can thus be rewritten as follows:

$$\max_{\mathcal{P}_1,\ldots,\mathcal{P}_k} \ln\left[\sum_{a\in B_k} e^{\overline{v}_a + \lambda\ln(\mathcal{P}_a)}\right]$$

under the constraints that $\sum_{a\in B_k}\mathcal{P}_a = 1$, and $\mathcal{P}_a \geq 0$, $\forall a \in B_k$. Exponentiating the objective function, the Lagrangian of is given by:

$$\mathcal{L}(\mathcal{P}) = \sum_{a\in B_k} \mathcal{P}_a^{\lambda} e^{\overline{v}_a} - \psi\left(\sum_{a\in B_k}\mathcal{P}_a - 1\right) + \sum_{a\in B_k}\phi_a\mathcal{P}_a$$

The complementary slackness condition is $\phi_a P_a^0 = 0$ with $\phi_a \geq 0$. The first order condition with respect to $\mathcal{P}_h$ is:

$$\lambda(\mathcal{P}_h^{B_k})^{\lambda-1}e^{\overline{v}_h} - \psi + \phi_h = 0$$

As we have restricted the alternatives so that $P_h > 0$, $\forall h \in B_k$, the first order condition can be simplified to:

$$\mathcal{P}_h^{B_k} = \left(\frac{\psi}{\lambda}e^{-\overline{v}_h}\right)^{\frac{1}{\lambda-1}}$$

Summing over alternatives:

$$\sum_{a\in B_k}\mathcal{P}_a^{B_k} = \sum_{a\in B_k}\left(\frac{\psi}{\lambda}e^{-\overline{v}_a}\right)^{\frac{1}{\lambda-1}} = 1$$

This can be rewritten as:

$$\psi^{\frac{1}{\lambda-1}} \sum_{a \in B_k} \left(\frac{e^{-\overline{v}_a}}{\lambda} \right)^{\frac{1}{\lambda-1}} = 1$$

Thus the Lagrangian multiplier ψ is equal to:

$$\psi = \left[\frac{1}{\sum_{a \in B_k} \left(\frac{e^{-\overline{v}_a}}{\lambda} \right)^{\frac{1}{\lambda-1}}} \right]^{\lambda-1}$$

Replacing this value of the Lagrangian multiplier in the expression for $\mathcal{P}_h^{B_k}$:

$$\begin{aligned} \mathcal{P}_h^{B_k} &= \frac{\left(\frac{e^{-\overline{v}_h}}{\lambda} \right)^{\frac{1}{\lambda-1}}}{\sum_{a \in B_k} \left(\frac{e^{-\overline{v}_a}}{\lambda} \right)^{\frac{1}{\lambda-1}}} \\ &= \frac{e^{\overline{v}_h/(1-\lambda)}}{\sum_{a \in B_k} e^{\overline{v}_a/(1-\lambda)}} \end{aligned} \tag{A.2}$$

A.3 Optimal consideration set

If we plug in the expression for the optimal unconditional choice probabilities in (A.2) into the objective function in (A.1), we obtain the expected value from optimally choosing from an arbitrary set B_k:

$$\begin{aligned} \mathbb{E}_{B_k} &= \ln \left[\sum_{a=1}^{k} e^{\overline{v}_a + \lambda \ln\left(\mathcal{P}_a^{B_k}\right)} \right] \\ &= \ln \left[\sum_{a=1}^{k} e^{\overline{v}_a} (\mathcal{P}_a^{B_k})^{\lambda} \right] \\ &= \ln \left[\frac{\sum_{a=1}^{k} e^{\overline{v}_a/(1-\lambda)}}{\left(\sum_{l=1}^{k} e^{\overline{v}_l/(1-\lambda)} \right)^{\lambda}} \right] \end{aligned} \tag{A.3}$$

We have that $\mathbb{E}_{B_{k+1}} > \mathbb{E}_{B_k}$, for $k \leq N-1$, if and only if:

$$\frac{\sum_{a=1}^{k+1} e^{\overline{v}_a/(1-\lambda)}}{\left(\sum_{l=1}^{k+1} e^{\overline{v}_l/(1-\lambda)} \right)^{\lambda}} > \frac{\sum_{a=1}^{k} e^{\overline{v}_a/(1-\lambda)}}{\left(\sum_{l=1}^{k} e^{\overline{v}_l/(1-\lambda)} \right)^{\lambda}}$$

Moving terms around:

$$\frac{\sum_{a=1}^{k+1} e^{\overline{v}_a/(1-\lambda)}}{\sum_{a=1}^{k} e^{\overline{v}_a/(1-\lambda)}} > \left(\frac{\sum_{l=1}^{k+1} e^{\overline{v}_l/(1-\lambda)}}{\sum_{l=1}^{k} e^{\overline{v}_l/(1-\lambda)}}\right)^{\lambda}$$

which is always satisfied as the ratio that appears on both sides is always greater than 1, and $\lambda \in (0,1)$. Hence, $\mathbb{E}_{B_{k+1}} > \mathbb{E}_{B_k}$, $\forall k \leq N-1$, and the consideration set is thus $B_N = A$.

A.4 Analytical results of $C_1(\mathcal{P}_1)$ and $C_1(\mathcal{P}_1)/C(\mathcal{P}_1)$

If the take the partial derivative of $C_1(\mathcal{P}_1)$ with respect to $\mathcal{P}_1$, we obtain:

$$\begin{aligned}\frac{\partial C_1(\mathcal{P}_1)}{\partial \mathcal{P}_1} &= -\lambda\left[\ln \mathcal{P}_1 + 1\right] + \lambda \int_{-\infty}^{+\infty} \frac{\partial \mathcal{P}_1(x)}{\partial \mathcal{P}_1}\left(\ln \mathcal{P}_1(x) + 1\right) g(x)\,\mathrm{d}x \\ &= -\lambda\left[\ln \mathcal{P}_1 + 1\right] + \\ &\quad + \frac{1}{\mathcal{P}_1(1-\mathcal{P}_1)}\left[\int_{-\infty}^{+\infty}\left[\ln \mathcal{P}_1(x) + 1\right]\mathcal{P}_1(x)[1-\mathcal{P}_1(x)]g(x)\,\mathrm{d}x\right]\end{aligned} \tag{A.4}$$

as:

$$\frac{\partial \mathcal{P}_1(x)}{\partial \mathcal{P}_1} = \frac{1}{\lambda}\frac{\mathcal{P}_1(x)}{\mathcal{P}_1}\frac{1-\mathcal{P}_1(x)}{1-\mathcal{P}_1}$$

When alternatives are *ex ante* identical, i.e., $\mathcal{P}_1 = 1/2$, we have that:

$$\mathcal{P}_1(x) = \frac{1}{1+e^{-x/\lambda}}$$

We can thus rewrite (A.4) as follows:

$$\frac{\partial C_1(\mathcal{P}_1)}{\partial \mathcal{P}_1}\Big|_{\mathcal{P}_1=1/2} = \lambda\left[\ln(2) - 1\right] - 4\int_{-\infty}^{+\infty} k(x)h(x)\,\mathrm{d}x \tag{A.5}$$

where we have defined:

$$h(x) \equiv \frac{1}{1+e^{-x/\lambda}}\left(1 - \frac{1}{1+e^{-x/\lambda}}\right) g(x)$$

and:

$$k(x) \equiv \ln(1+e^{-x/\lambda}) - 1$$

As the function $h(x)$ is symmetric around zero, i.e., $h(x) = h(-x)$, while the function $k(z)$ is such that $k(x)+k(-x) \geq 0$, we can conclude that the integral appearing in (A.5) is positive, and thus:

$$\frac{\partial C_1(\mathcal{P}_1)}{\partial \mathcal{P}_1}\Big|_{\mathcal{P}_1=1/2} < \lambda\left[\ln(2) - 1\right] < 0 \tag{A.6}$$

This also implies that, when $\mathcal{P}_1 = 1/2$, $\partial C_2(\mathcal{P}_1)/\partial \mathcal{P}_1 > 0$, $\partial C(\mathcal{P}_1)/\partial \mathcal{P}_1 = 0$, and the share of the total cost devoted to alternative 1 is decreasing.[51]

B Full-information RUM model

B.1 Unobserved heterogeneity and $\mathcal{P}_1$

Consider a full-information RUM model with unobserved heterogeneity describing the location-decision problem that migrants from a given origin face.[52] Let $v_k = \alpha\left(\ln y_k - \ln \tau_k\right)$ represent the deterministic component of utility associated with migrating to k, with k denoting one of the N alternatives belonging to the choice set A. Let us introduce the canonical assumption that the individual-specific stochastic component of utility ϵ_{ik} is i.i.d. EVT-1, with a scale parameter $\sigma > 0$. The variance of this distribution is equal to $(\pi^2/6)\sigma^2$, so that a greater value of σ reflects a greater unobserved heterogeneity in location-specific utility. The probability that a migrant finds optimal to opt for destination $k \in A$ is given by (McFadden, 1978):

$$\mathcal{P}_k = \frac{e^{v_k/\sigma}}{\sum_{a\in A} e^{v_a/\sigma}} \tag{B.1}$$

A key property of this discrete choice model is the independence from irrelevant alternatives, i.e., $\ln\left(P_k/P_l\right) = (v_k - v_l)/\sigma$, $\forall k,\, l \in A$. An implication of this fundamental property is that the marginal effect of a variation in the deterministic component of utility on the log odds ratio $\ln\left(P_k/P_l\right)$ is independent from $\mathcal{P}_k$ and $\mathcal{P}_l$, i.e., from v_k and v_l. The partial derivative of $\ln \mathcal{P}_k$ in (B.1) with respect to v_k is given by:

$$\frac{\partial \ln \mathcal{P}_k}{\partial v_k} = \frac{1}{\sigma}(1 - \mathcal{P}_k) \geq 0 \tag{B.2}$$

[51] Following the same steps, we can also extend the results about the slope of $C_1(\mathcal{P}_1)$ to any $\mathcal{P}_1 \geq 1/2$.

[52] We avoid, as in Section 2 introducing origin and time subscripts to avoid cluttering the notation.

Without loss of generality, let us assume that $v_1 \geq v_2 \geq ... \geq v_N$, so that $P_1 \geq P_k$, $\forall k \in A \setminus \{1\}$. If we compute the partial derivative of $\ln P_1$ with respect to σ, we obtain:

$$\frac{\partial \ln \mathcal{P}_1}{\partial \sigma} = -\frac{1}{\sigma^2}\left(v_1 - \sum_{a \in A} v_a \mathcal{P}_a\right) \leq 0 \tag{B.3}$$

with the inequality in (B.3) holding strictly whenever $\mathcal{P}_1 > 1/N$. Thus, when σ is lower, then the probability $\mathcal{P}_1$ of opting for the alternative that is, on average, most attractive increases, and the responsiveness of the choice probabilities $\mathcal{P}_k$ with respect to variations in the deterministic component of utility v_k gets magnified. This, in turn, implies that even in a full-information RUM model the share of migration flows in the main destination is correlated with the size of the estimated coefficients, but in a way that is opposite to the one that characterizes a model with costly information acquisition.