JUE Insight: Moving Cost Magnitudes in Moving Cost Models

Greg Howard*

August 29, 2025

Abstract

The internal migration literature has estimated a wide range of moving costs, including some that are several times larger than annual income. How should economists interpret this estimate? I show that in standard models, average moving costs can be decomposed into an "information" term and a "returns to migration" term. The information term is proportional to the Shannon entropy of next period's location minus the Shannon information of staying in the same location. In simple models, the information term is much larger than the returns to migration term; in some cases, the returns to migration term is zero. Therefore, average moving costs are a helpful statistic about the model's predictive power regarding future moves but are not invariant to seemingly innocuous choices of the modeler.

JEL Codes: R23, D80, J61, F16

Keywords: internal migration, labor mobility, information costs, Shannon entropy

*University of Illinois, Urbana-Champaign. Thanks to Treb Allen, Vivek Bhattacharya, Jonathan Dingel, Jorge Lemus, Charly Porcher, Tyler Ransom, and participants at Midwest Macro, the Urban Economics Association meetings, and the Illinois macro lunch for helpful comments and to Flavio Rodrigues for research assistance. All errors are my own.

Papers in the literature on internal migration have generated a wide range of moving cost estimates, some of which are many times the average annual income (Kennan and Walker, 2011; Bryan and Morten, 2019, etc.). Others have noted that these migration costs could reflect other frictions, and that if they explicitly model these other frictions, then estimated moving costs fall (Schmutz and Sidibé, 2019; Porcher, 2020; Heise and Porzio, 2022; Giannone, Li, Paixao and Pang, 2023).[1]

Jia, Molloy, Smith and Wozniak (2023), a review article in the *Journal of Economic Literature*, summarizes the state of the literature as, "while unobserved and potentially very large costs might help explain migration rates that are low relative to the potential earnings gains from migration, different models imply substantively different estimates of the size of these costs."[2]

In this paper, I propose a different way to think about these estimated moving costs. I show that average moving costs can be decomposed into two terms. The first is an "information" term, which is proportional to the average Shannon entropy of next period's location minus the Shannon information of next period's location being the same as the current location (Shannon, 1948). The second term measures the "average returns to migration," which in many models is quantitatively small. Shannon information is a measure of how surprising an event is: the more unlikely it is to happen, the more information it contains. Shannon entropy measures expected information before the realization of the event. Thus, moving costs primarily measure how surprised the modeler will be to learn an individual's next location, relative to the unsurprising outcome that the individual stayed put.[3]

[1]Many papers use moving cost models but do not report average moving costs, perhaps because the authors find them implausibly large. For example, the model in Caliendo, Dvorkin and Parro (2019) would imply large moving costs, but they develop solution techniques that do not require backing out the moving cost parameters. Another example is Schubert (2021), which does not report the average moving cost, but does consider counterfactuals in which the moving costs change.

[2]Other methodologies of uncovering migration costs give various results. Koşar, Ransom and van der Klaauw (2022) uses a survey to the estimate the willingness to pay to avoid moving, and estimates an average moving cost of $54,000.

[3]The modeler is not going to be surprised by aggregate migration flows, which they will be able to match exactly in the data. Rather, this notion of surprise is for an individual's

Based on this result, I show that estimated moving costs change depending on the time period or the geographic partition of the model. Additionally, average costs are also sensitive to the modeler's information set regarding the agents. For example, knowing the birthplace of each person leads the modeler to estimate smaller moving costs. I give examples of the ways these modeling decisions affect average moving costs using data from the 2000 Census and the American Community Survey.

However, that does not mean moving costs are uninteresting. In particular, comparing moving costs across models is informative of how good those models are at predicting an individual agent's next period location. This alternative interpretation makes sense of some recent results, specifically that richer models of moving—which typically incorporate more information—exhibit smaller moving costs.

What is it about standard models that leads to this relationship between moving costs and Shannon information? The critical assumption is the i.i.d. extreme value shocks. The specific functional form is critical for generating the exact Shannon entropy term. More importantly, the i.i.d. assumption is what generates the dependence of estimated moving costs on the timing and geography choices of the modeler. When the modeler makes these choices, they are also making an explicit assumption of how much and how often agents are given opportunities to move. A simple way to see this is that, absent any moving costs, when agents draw new shocks more often or for more locations, they will move more. Hence, moving costs have to be higher to match the same rate of migration in the data.[4]

This paper speaks to the literature on estimates of migration costs, which I discuss in detail in Section 3, as well as the quantitative migration literature more generally. One implication of this paper is that it is not a helpful exercise to compare estimated moving costs to one's priors on average moving costs or the potential gains of migration. If one thinks that moving costs are "too

location choice, which depends on the realization of a random shock.

[4]The sensitivity of moving costs to these choices has been recognized in the literature (e.g. footnote 12 from Kennan and Walker (2011)), but formalizing it via Shannon entropy is novel.

high," there are somewhat arbitrary changes one could make to the model to reduce them: using longer time periods between utility shocks, using fewer locations, or increasing the information in the model. Many papers in the literature that report a moving cost include some explanation for why their estimates might be feasible;[5] this is not a useful way to judge the applicability of the model. Other papers that estimate relatively low moving costs will highlight it as a feature of their model;[6] while it is true that models with richer information will feature lower moving costs, it is not true that having lower moving costs makes a model more realistic *per se*.

Importantly, my results on the sensitivity of migration costs to modelers' arbitrary choices mostly do not extend to the effects of a regional shock or to policy counterfactuals, which are some of the main focuses of that literature. The most common exception would be counterfactuals which reduce moving costs by a percentage of the initial cost; since this counterfactual is based on the measurement of moving costs, the results of this counterfactual will also be sensitive to arbitrary choices of the modeler.

This paper has similarities to the literature that relates discrete choice models to entropy. Wilson (1969) sets up a problem of entropy maximization to model transportation choice, showing that the choices resemble a gravity equation, and Anas (1983) shows this is the same problem as a typical random utility discrete choice problem, as we study here. More recent work has established an equivalence between random utility discrete choice problems and rational inattention, where agents pay a cost related to the entropy reduction of the information they acquire (Matějka and McKay, 2015; Fosgerau, Melo, De Palma and Shum, 2020). To my knowledge, no one has related estimated average moving costs to entropy, as I do here.[7]

[5]For example, Kennan and Walker (2011) works through an example of an agent whose returns to migration are of approximately the same magnitude as the average cost that they estimate. They suggest this makes their estimate more plausible.

[6]Heise and Porzio (2022) say "[Our] relatively small moving costs and home biases reflect our model's ability to separate their impact from spatial search frictions and from general labor market frictions."

[7]Porcher (2020) and Bertoli, Moraga and Guichard (2020) are perhaps the closest papers, because they involve both Shannon entropy and migration. Those papers assume rationally

The steady-state result of this paper could also be extended to the international trade literature, where it would relate average trade costs to the Shannon entropy of a good's destination minus the Shannon information of it being consumed at home. This would hold when trade followed a gravity equation and was balanced (the analog of the steady-state assumption in this paper). I do not focus on this application because average trade costs are not commonly reported.[8]

1 Standard moving cost model

In this section, I use the standard moving cost model to derive an interpretable expression for average moving costs. In this model, agents choose their location to maximize the present value of their utility. As part of that utility, they face moving costs and draw independent and identically distributed (i.i.d.) extreme value shocks for every location in every time period.[9]

There is a continuum of people indexed by n that live in discrete locations indexed by i and who have state variables indexed by s. Time is discrete and indexed by t. The population of people living in i with state variables s at time t is denoted by $p_{it}(s)$. The share of people in i who move from i to j at time t is denoted $m_{i\to j,t}$. Based on this notation, $m_{i\to i,t}$ will refer to the non-migration rate in i. m_{it} denotes the total outmigration share from i to all locations $j \neq i$ at time t. When referring to steady-states, the t index is

inattentive agents, and a typical assumption for rational inattention is that the costs that agents have to pay is related to the Shannon entropy of the information they acquire, as in Matějka and McKay (2015). However, the critical difference is that this paper emphasizes the moving costs as a measure of the *modeler's* lack of information, whereas those paper emphasizes that *agents'* lack of information can resemble moving costs.

[8]My interpretation may be helpful in the literature that estimates workers' switching costs across industries, as in Dix-Carneiro (2014), which estimates switching costs to be greater than annual income.

[9]Some versions of the standard model, including Kennan and Walker (2011), assume the i.i.d. extreme value shocks are part of the moving costs. When including the shocks as part of moving costs, estimated average moving costs are negative. However, their most well-known statistic does not include the shocks as part of the moving costs: "For the average mover, the cost is about $312,000 (in 2010 dollars) if the payoff shocks are ignored" (Kennan and Walker, 2011, p. 232).

dropped. Moving costs are bilateral between two locations, so δ_{ij} refers to the moving cost from i to j. I assume there is no cost to not moving, i.e. $\delta_{ii} = 0$ for all i. I use the notation $\mathbb{E}_{is}$ to refer to the population-weighted average across i and s and $\mathbb{E}^m$ to refer to the migration-weighted average. I will be interested in the average migration cost, which we define to be $\bar{\delta} \equiv \mathbb{E}^m[\delta_{ij}]$.

Each agent has a state variable s that may affect their moving cost and utility, and can evolve arbitrarily. In addition, their current location i is a state variable, which we will write separately. Agents maximize the following value function:

$$V_{nt}(i, s) = \log w_{it}(s) + a_{it}(s) + \max_j \left\{ -\delta_{ij}(s) + \frac{1}{\mu}\epsilon_{jnt} + \beta \mathbb{E} V_{nt+1}(j, s'(i, s, j, X)) \right\}$$

where $w_{it}(s)$ is the real wage, $a_{it}(s)$ is the amenities in i, $\delta_{ij}(s)$ is the moving cost from i to j, and ϵ_{jnt} is an i.i.d. type 1 extreme value shock. μ is a scale parameter, which governs the elasticity of substitution between places. s' is a function of the existing states i and s, the choice variable j, and a random variable X.

Define $v_{jt}(i, s) \equiv \mathbb{E} V_{nt+1}(j, s'(i, s, j, X))$. Then migration is given by

$$m_{i \to j,t}(s) = \frac{\exp(\mu(\beta v_{jt}(i, s) - \delta_{ij}(s)))}{\sum_k \exp(\mu(\beta v_{kt}(i, s) - \delta_{ik}(s)))}$$

Because $\delta_{ii}(s) = 0$, $\delta_{ij}(s)$ can be written

$$\delta_{ij}(s) = \beta v_{jt}(i, s) - \beta v_{it}(i, s) - \frac{1}{\mu} \log m_{i \to j,t}(s) + \frac{1}{\mu} \log m_{i \to i,t}(s) \tag{1}$$

Consider the migration-weighted average moving cost, which is often reported in papers in the literature:

$$\bar{\delta} \equiv \mathbb{E}^m[\delta_{ij}] \equiv \frac{\sum_{s,i,j:i \neq j} p_i(s) m_{i \to j}(s) \delta_{ij}(s)}{\sum_{s,i,j:i \neq j} p_i(s) m_{i \to j}(s)} \tag{2}$$

The main proposition relates $\bar{\delta}$ to measures of information about future locations. Before stating the proposition, it is helpful to define some additional notation.

Define J to be a discrete random variable: next period's location. Lowercase j will refer to specific realizations of J. I use the notation $H(J|i)$ to refer to the Shannon entropy of J for a person currently living in i, and the notation $I(j|i)$ to refer to the Shannon information of the realization of $J = j$ given i, i.e. migrating from i to j. Since $m_{i \to j}$ is the migration probability for someone living in i to move to j,

$$I(j|i) = -\log m_{i \to j}$$

and

$$H(J|i) = -\sum_j m_{i \to j} \log m_{i \to j}$$

based on the mathematical definitions of Shannon information and entropy (Shannon, 1948).

An informal way to understand Shannon information is that it measures how surprising an event is. Since most people do not move, the event of not moving is unsurprising, and the Shannon information of not moving is small. Shannon entropy measures the expected Shannon information. So if it is hard to predict where people will live next period, then the Shannon entropy will be large. Another way to think about Shannon entropy is that Shannon entropy is approximately proportional to the number of "yes or no" questions one would have to ask in order to acquire the information, i.e. the number of bits the information contains.[10] So $H(J|i)$ is proportional to the bits of information needed to communicate where a person in i will live next period.

Note that in the migration context, Shannon entropy depends on the modeler's choice of time and location: it is easier to predict locations in the next period if the length of a period is short instead of long or if the geography is coarse, like U.S. states, instead of fine, like U.S. counties. This will be an important feature for applications of the main proposition.

[10] This is an approximation because Shannon entropy is a continuous measure. It can be scaled by $\log 2$ to convert the units of Shannon entropy into bits.

Proposition 1. *The average moving cost in a moving cost model can be decomposed into two parts. The first measures relative information: it is the average Shannon entropy of next period's location minus the Shannon information of not moving, all divided by the average moving rate times the migration elasticity. The second is the average returns to migration. In math,*

$$\bar{\delta} = \underbrace{\frac{1}{\mu}\frac{1}{\mathbb{E}_{is}m_{is}}\mathbb{E}_{is}[H(J|i,s) - I(i|i,s)]}_{\textit{Information Term}} + \underbrace{\beta\mathbb{E}^m[v_{jt}(i,s) - v_{it}(i,s)]}_{\textit{Returns to migration}} \quad (3)$$

The information term can alternatively be written

$$\bar{\delta} = \underbrace{\frac{1}{\mu}\mathbb{E}^m\left[H(J|i \to \not{i}, s) + I(\not{i}|i,s) - I(i|i,s)\right]}_{\textit{Information Term}} + \underbrace{\beta\mathbb{E}^m[v_{jt}(i,s) - v_{it}(i,s)]}_{\textit{Returns to migration}} \quad (4)$$

where $H(J|i \to \not{i}, s)$ is the Shannon entropy of next period's location conditional on moving, $I(\not{i}|i,s)$ is the Shannon information of moving and $I(i|i,s)$ is the Shannon information of not moving.

Proofs are collected in Appendix A. The second formulation is in some contexts helpful because it separates out the Shannon entropy conditional on moving from the information involved in moving or not.[11]

In simple applications, the information term is significantly larger than the returns to migration. In fact, in the following two corollaries, I present two special cases in which the returns to migration term disappears entirely because the returns of people moving from i to j are canceled out by the

[11]For example, I can extend Proposition 2 to ϵ shocks that are nested logit as in Monras (2020), where there is one elasticity for choosing whether to move at all and one elasticity for choosing which location to move to. Assuming no state variables s, the formula becomes

$$\bar{\delta} = \mathbb{E}^m\left[\frac{1}{\mu}H(J|i \to \not{i}) + \frac{1}{\lambda}(I(\not{i}|i) - I(i|i))\right] + \text{returns to migration}$$

where μ is the migration elasticity across destination locations and λ is the migration elasticity of moving at all. This is intuitive given that the I terms are about the information of whether to move at all, and the H term is about the information conditional on moving. See Appendix B for details.

people moving from j to i. The first corollary considers cases in which the state variables (except i) are fixed types, such as race or birthplace, and the model is in steady-state. It nests the version in which the only state variable is i.

Corollary 1. *Suppose that the state variables are fixed, i.e.* $s' = s$, *and that the model is in steady-state, i.e. the population shares of type* s *in location* i *are not changing between* t *and* $t+1$. *Then, the average moving cost is the average Shannon entropy of next period's location minus the Shannon information of not moving, all divided by the average moving rate times the migration elasticity. In math,*

$$\bar{\delta} = \frac{1}{\mathbb{E}_{is} m_{is}} \frac{1}{\mu} \mathbb{E}_{is}[H(J|i,s) - I(i|i,s)] \tag{5}$$

The second corollary considers state variables that are independent of past state variables. This would include models in which decisions are made sequentially, such as agents draw a random consideration set (which would be considered the state variable) and then maximize among choices in that set. It would also cover cases of receiving job offers that depend only on the current location, or acquiring information as long as that information does not depend on previous state variables, such as the model in Porcher (2020).

Corollary 2. *Suppose that the future state variables are independent of current state variables, i.e.* s' *is a function of only* j *and* X, *and that the model is in steady-state, i.e. the population in location* i *is not changing between* t *and* $t+1$. *Then, the average moving cost is the average Shannon entropy of next period's location minus the Shannon information of not moving, all divided by the average moving rate times the migration elasticity. In math,*

$$\bar{\delta} = \frac{1}{\mathbb{E}_{is} m_{is}} \frac{1}{\mu} \mathbb{E}_{is}[H(J|i,s) - I(i|i,s)] \tag{6}$$

Of course, typical models in the literature will not have the returns to migration term cancel out exactly as in these corollaries. Our next proposition shows one way to potentially approximate the "returns to migration" term outside of steady-state.

Proposition 2. *Suppose state variables are fixed, i.e. $s' = s$. If outmigration-weighted average moving costs into and out of every location are equal, in the sense that*

$$\sum_i m_{i\to j}(s)\delta_{ij}(s) = \sum_i m_{i\to j}(s)\delta_{ji}(s) \tag{7}$$

then the "returns to migration" term is given by

$$\beta\mathbb{E}^m[v_j(s) - v_i(s)] = \frac{1}{2\mu}\mathbb{E}^m \log\left(\frac{m_{i\to j}(s)}{m_{j\to i}(s)}\frac{m_{j\to j}(s)}{m_{i\to i}(s)}\right)$$

In this formula, the relative migration from i to j versus j to i—relative to the non-migration rates—is being used to infer how much better the destination location is than the origin location. While the equality of average moving costs is unusual, having purely symmetric moving costs, $\delta_{ij} = \delta_{ji}$, will imply condition (7); but this condition still allows the model to exactly match the data, which pure symmetry does not. Nonetheless, in many applications, symmetry of migration costs is assumed, as in the Head and Ries (2001) index and in Bryan and Morten (2019), or when migration costs are projected onto distance, as in Kennan and Walker (2011), or any other metric (e.g. linguistic distance as in Wang (2024)). In Section 2.1, we will plug in observable migration rates into this formula to compare them to the size of the information term.

2 Moving costs sensitivity

In this section, I present three observations related to the main proposition that highlight the sensitivity of migration costs to a modeler's choices. This section primarily offers a new perspective on existing knowledge. It has long been recognized that moving costs are sensitive to the model used to estimate them (e.g. Kennan and Walker, 2011), so the contribution here is to show that the formulation as information can give a helpful perspective on this sensitivity.

Observation 1. *Holding the migration elasticity constant, the information term depends on the modeler's choice of length of the time period.*

One way to see this observation is from the second formula in Proposition 1.[12] Over short time horizons, the Shannon entropy conditional on moving, $H(J|i \to \not{i})$, does not vary much. However, migration rates are smaller for shorter time horizons. So based on Proposition 1, average moving costs will vary with the time period chosen as $I(\not{i}|i) - I(i|i) = \log \frac{1-m_i}{m_i}$ increases when time horizons are short. In fact, as time horizons get arbitrarily short, estimated average moving costs get arbitrarily large.

Observation 2. *Holding the migration elasticity constant, the information terms depend on the modeler's choice of geographic partition.*

The Shannon entropy of next period's location depends on how the modeler partitions geography.[13] Generally, the more locations there are, the harder it is to predict exactly which one any given person will end up in. Therefore, one would expect that Shannon entropy would increase in the number of locations.[14] Mechanically, migration rates also increase in the number of locations. To my knowledge, there is no way to order geographies such that estimated migration costs must increase or decrease, but in the empirical results, I show that the Shannon entropy change dominates the change in the information of not moving when I apply it to states versus migration public use microdata areas (MIGPUMAs). Certainly, there is no reason to expect the change in Shannon entropy and the change in migration rates to cancel out.

[12] Alternative intuition for this observation can be seen directly in equation (3). The migration rate $m_{i\to j}$ is increasing in the time horizon, and the non-migration rate $m_{i\to i}$ is decreasing in the time horizon, so if $v_{it} - v_{jt}$ is not changing with the time horizon, estimated migration costs must decline.

[13] Again, equation (3) gives some hints at this proposition because if we divide a region into two regions, the migration rate to either individual region will be less than to the original region. The model estimates higher moving costs to rationalize these lower migration rates. This point is acknowledged in Kennan and Walker (2011) but the quantitative implications are not explored.

[14] This may not be true in all cases, if the more precise location is sufficiently informative of future locations.

Consider two extreme examples to show that moving costs can be estimated to be large or small depending on the partition. We will use the second formula in Proposition 1 for these examples and assume for the purposes of this exercise that i is the only state variable. In the first case, consider partitioning every house into its own geography. In the 2000 Census, 43 percent of people moved houses in the previous 5 years. The Shannon entropy conditional on moving is enormous because it is almost impossible to predict the exact house that anyone would live in. So based on equation (4), we would have an enormous number plus $\log((1-0.43)/0.43)$. Just to put a number on it, we can assume that modeler can assign no individual house a probability of being chosen of greater than 0.1 percent. Then a lower bound for the information term is

$$\frac{1}{\mu}\mathbb{E}^m[H(J|i \to \not{i}) - I(\not{i}|i) - I(i|i)] \leq \frac{1}{\mu}\left(-\log\frac{1}{1000} + \log\frac{1-0.43}{0.43}\right) \approx \frac{7.2}{\mu}$$

Alternatively, we could partition the United States into houses with an even-numbered address and ones with an odd-numbered address. If we assume it is random which type of house you move into, we would expect 21.5 percent of the population to "move regions." Conditional on "moving regions," the Shannon entropy is zero. So the information term is

$$\frac{1}{\mu}\mathbb{E}^m[H(J|i \to \not{i}) - I(\not{i}|i) - I(i|i)] = \frac{1}{\mu}\left(0 + \log\frac{1-0.215}{0.215}\right) \approx \frac{1.3}{\mu}$$

Stepping back from the model, this partition should not matter. The "true" average cost of moving from an even-numbered house to an odd-numbered house should not be different than the "true" average cost of moving between any two houses. Yet, how we partitioned the geography changed the *estimated* information part of the moving cost by a factor of more than 5.[15]

[15] A reader might object by stating that the correct way to interpret the lower number is that it is the cost of moving relative to the cost of "not moving," and that "not moving" includes many people who do change houses. But under that interpretation, I can use the estimate of moving costs from the first model and compare the cost of moving ($7.2/\mu$) to the weighted average of moving costs for people who remain in same-parity house: $(0.5 \times .43 \times 7.2/\mu + .57 \times 0)/(0.5 \times .43 + .57)$. This difference is approximately $5.2/\mu$, which is still about four times as large compared to $1.3/\mu$. So this alternative interpretation cannot

Observation 3. *Holding the migration elasticity constant, the information term depends on the modeler's information set.*

Suppose the modeler knew some immutable characteristic of individuals, s, such as their race or their birthplace. If they estimate separate moving costs by this characteristic, then the information term is:

$$\frac{1}{\mathbb{E}_{is}m_{is}}\frac{1}{\mu}\mathbb{E}_{is}\left[H(J|i,s)-I(i|i,s)\right] \tag{8}$$

But if they do not observe this characteristic, then

$$\frac{1}{\mathbb{E}_{i}m_{i}}\frac{1}{\mu}\mathbb{E}_{i}\left[H(J|i)-I(i|i)\right] \tag{9}$$

Shannon entropy is concave, and Shannon information is convex, so by Jensen's inequality, $\mathbb{E}_{is}[H(J|i,s)-I(i|i,s)] \leq \mathbb{E}_{i}H(J|i)-I(i|i)]$. Since $\mathbb{E}_{is}m_{is}=\mathbb{E}_{i}m_{i}$, then the expression in (8) is weakly smaller than (9). If s provides any information about the next periods' location, then the inequality is strict.

This expression also holds for state variables permitted in Corollary 2. For example, if the modeler modeled the decision making process in two stages where, first, each person chooses a consideration set, and second, compares the utilities available in each, s could be the consideration set. In this setup, the returns to migration term will still drop out in steady-state. So in a model with consideration sets, the modeler will estimate lower moving costs than in a model without consideration sets.

A key assumption of the previous corollaries was holding μ constant. Without this assumption, migration costs could be constant across these choices if μ changed instead. However, this may be undesirable because μ represents the elasticity of migration to wages and is central to many counterfactual questions. In other words, changing μ changes the shock propagation and policy counterfactuals of the model, which are otherwise similar across choices of timing and geography. Nonetheless, an alternative weaker interpretation of these corollaries is that *the product of migration costs and migration elasticity*

reconcile these estimates.

is sensitive to arbitrary choices of the modeler.

2.1 Moving cost calibrations with data

In this section, I illustrate the observations from the previous section using real world data. In particular, I estimate the average moving costs using equation (3) with data from the Census and the American Community Survey (ACS) in 2000 (Ruggles, Flood, Sobek, Brockman, Cooper, Richards and Schouweiler, 2023).[16] For each state-pair, I calculate $m_{i \to j}$ as the share of people who lived in state i that moved to state j, either from 1995 to 2000 in the Census, or from 1999 to 2000 in the ACS. I also calculate $m_{i \to j,b}$, where I calculate the probability of moving from i to j given a birthplace b. And I also calculate $m_{i \to j}$ where i and j are Migration Public Use Microdata Areas (MIGPUMAs)—a within-state region with at least 100,000 people—instead of states.

Kennan and Walker (2011) and many subsequent papers express moving costs in dollar terms. Since wages enter utility in logs, one can interpret these average moving costs as a percent of wages.[17] Therefore, one might think of moving costs as a measure of the expected Shannon information minus the Shannon information of not moving, where each bit of information "costs" $\frac{w}{\mu \log 2}$ dollars per migrant.[18]

I then calibrate the average moving costs according to equation (3) and Proposition 2. In the literature, there is little consensus on what μ is, and some good arguments that typical methods have not estimated it well (Borusyak, Dix-Carneiro and Kovak, 2022), so I use $\mu = 1$ because it is easy for the reader to scale the moving costs by whatever μ they prefer.[19] For the dollar

[16]This is the only year, to my knowledge, in which similar surveys asked about the 1-year migration rate (the ACS) and the 5-year migration rate (the Census).

[17]Kennan and Walker (2011) actually expresses utility in dollar terms directly, so there is no need for this adjustment. However, much of the subsequent literature does express wages in logs.

[18]When we change the time period to five years, the most natural change to the model is to change $\log w_{jt}$ in the value function to $5 \log w_{jt}$ to minimize changes to the level of utility or the marginal utility. In this case, the moving cost can still be interpreted as a percent of annual wages. If I changed the utility term to $\log(5w_{jt})$, the change in average moving costs would be more dramatic.

[19]Borusyak et al. (2022) makes the point that regressing the change in population on labor

Table 1: Estimated Moving Costs for Different Models

	(1) Migration Rate	(2) Information Term	(3) Returns to Migration	(4) Estimated Moving Cost	(5) Cost in $1000's	(6) Cost in $1000's ($\mu = 1/3$)
1 year, states	0.024	6.692	0.104	6.796	319	958
	(0.0002)	(0.0150)	(0.0097)	(0.0159)	(0.75)	(2.24)
5 year, states	0.085	5.585	0.022	5.607	264	791
	(0.0001)	(0.0015)	(0.0007)	(0.0013)	(0.06)	(0.19)
5 year, states (modeler knows birthplace)	0.085	4.981	-0.260	4.722	222	666
	(0.0001)	(0.0019)	(0.0013)	(0.0014)	(0.07)	(0.20)
5 year, MIGPUMAs	0.173	5.983	0.055	6.038	284	851
	(0.0001)	(0.0015)	(0.0005)	(0.0015)	(0.07)	(0.21)

Notes: All datasets are from 2000. 1 year migration uses migration measured over 1 year from the ACS. 5 year migration uses migration measured over 5 years from the Census. The unit of geography is a state or a MIGPUMA, a subset of a state with at least 100,000 people in it. Birthplace is an indicator variable either for the state of birth or for being from anywhere outside the 51 U.S. states. Columns (2), (3), (4), and (5) assume $\mu = 1$. Column (2) uses the formula from Proposition 1. Column (3) uses the formula from Proposition 2, and column (4) is the sum of columns (2) and (3). The median household income in 2000 (for people also living in the U.S. in 1995) was $47,000, so column (5) is column (4) times 47. Standard errors, in parentheses, are bootstrapped with 1000 replications.

values, I also include a column for $\mu = \frac{1}{3}$, following Caliendo et al. (2019) and Kleinman, Liu and Redding (2023).

The first result of note is that under the assumptions of Proposition 2, the returns to migration term is much smaller than the information term, with the ratio of magnitudes ranging from 19 to 250. In fact, in these simple models, even outside of the assumptions of Proposition 2, the returns to migration would have to be—on average—at least the value of annual income to be on the same magnitude as the information term.[20]

The comparisons of results across rows are consistent with our 3 observations. In the 1-year calibration, I estimate moving costs of $6.7/\mu$ log points, or when converted to dollars, \$319,000 when $\mu = 1$ or \$958,000 when $\mu = \frac{1}{3}$. This is the same order of magnitude as Kennan and Walker (2011), who estimated moving costs of \$312,000 (p. 232).

In the 5 year calibration, I estimate smaller moving costs: $5.6/\mu$ log points, or \$264,000/μ.[21] This is because at the 5-year horizon, an individual choosing to move is less surprising than at the 1-year horizon.

If the modeler knows the birthplace, the entropy decreases since birthplace is a helpful predictor of future location choices. Compared to the 5 year calibration where the modeler does not know birthplace, the moving cost is even lower: $4.7/\mu$ log points (\$222,000/μ). This is consistent with Zerecero (2021).

demand shocks—even well-identified labor demand shocks—does not identify μ because the shocks are correlated across space and affect both origin and destination locations.

[20]Even outside of steady-state, the returns for most people moving from i to j is canceled out by someone moving from j to i. For example, in the "5 year, states" row, only a 0.0107 share of the gross migration leads to net-positive migration for the receiving state. For the returns to migration term to be even 10 percent of the size of the information term, the average returns for someone moving from a negative-net-migration state to a positive-net-migration state would have have to be

$$10\% \times \frac{1}{\mu} \times 5.85 \times \frac{1}{.0107} \times \$47{,}000 = \frac{\$2.6 \text{ million}}{\mu}$$

Even for $\mu = 1$, this is significantly larger than estimates of the returns to migration.

[21]This difference is because of the different time horizons, not the different datasets. Based on equation (4), adjusting only the gross migration probability in column (1), which would influence the Shannon information of moving and not moving, while keeping the Shannon entropy conditional on moving the same would account for slightly more of the difference in the information term in Column (2).

One implication is that if the true model of the world was the model described in Section 1 and if moving costs and utility depended on birthplace, but the modeler incorrectly estimates the model without accounting for birthplace, then they would estimate moving costs that are about 15 percent too high.

Finally, if I use MIGPUMAs instead of states, it is much harder to predict future locations, since MIGPUMAs are a finer geography. The moving costs increase by about $0.4/\mu$ when I use MIGPUMAs instead of states, to $6.0/\mu$ log points (\$284,000/μ). This means if the true model involved drawing an i.i.d. shock for every MIGPUMA, but the modeler mistakenly assumed the i.i.d. draws were for every state, they would underestimate moving costs by a bit less than 10 percent.[22]

3 How should moving costs be interpreted?

Proposition 1 gives a new interpretation of moving costs in the steady-state of the standard migration model. In this section, I show how that interpretation can help make sense of the literature's estimates of moving costs.

Before I discuss this application, it is important to address three ways in which moving costs differ from the information term in Proposition 1. First, in almost all examples in the literature, moving costs are estimated for models that are not in steady-state. Second, some models in the literature have state variables that depend on previous locations and state variables that do not fit into the assumptions of Corollaries 1 or 2. Third, the literature often estimates moving costs through gravity equations or the index from Head and Ries (2001), rather than matching migration flows exactly as in equation (3).

With the aforementioned state variables, we also have to account for the returns to migration that come from migration being associated with higher returns. The state variables that seem most likely to make these gains large are previous locations (because moving is sometimes assumed to lower future

[22]Note that they underestimate average moving costs even though their average is over only interstate moves, and more than half of the moves in the average of the true average moving cost are within-state moves.

moving costs) and employment status (if moving is associated with gaining a job). However, the gain from lower future migration costs has to be only a small fraction of the total moving costs.[23] And given that most unemployment spells have short durations, the gains of moving from unemployment to employment are likely only a fraction of annual income, whereas the information term is typically several times annual income, as in the previous section.

The final difference is that moving costs are often estimated using a gravity equation or the Head and Ries (2001) index.[24] However, equation (3) will hold within the model, so as long as the model's migration flows are approximately equal to the migration flows from the data, then the migration costs are approximately equal as well.

Given these three considerations, estimated migration costs are not going to be exactly given by the information term in Proposition 1. Nonetheless, the deviations should be quantitatively small.

So how should a reader interpret reported moving costs in an economics paper? I propose that they may want to compare the average moving costs to other papers or other model specifications, as I do in Table 2.[25] These

[23]Even among people that moved states the previous year, more than 70 percent will not move again the next year (Howard and Shao, 2025). Because non-migration is more common than migration, and because, by assumption the continuation value of moving is, on average, higher, then average migration costs are still positive for people that just moved. This means the average returns to migration are bounded above by the average migration costs by a standard envelope argument. But the average migration costs here include the zero-cost paid by the non-movers, which is the vast majority of people. So an upper bound on the returns to migration would be $\bar{\delta}$ times the migration rate. In other words, the returns to migration term will not account for more than 4 percent of the average migration cost.

[24]This formula assumes $\delta_{ij} = \delta_{ji}$. Then from equation (3),

$$\delta_{ij} = \frac{1}{2}\frac{1}{\mu}(\log m_{i \to i,t} + \log m_{j \to j,t} - \log m_{i \to j,t} - \log m_{j \to i,t})$$

[25]To include a paper in this table, I required the paper to report an average moving cost in some sort of interpretable units and to use extreme value shocks. Papers such as Bishop (2012) and Oswald (2019) report a moving cost function and seem to have moving costs in the same ballpark as Kennan and Walker (2011), but do not report average costs. Bartik and Rinz (2018) reports a moving cost of $683,000, but this is not the average of all movers; rather it is the average cost for a 500 mile move. Bayer and Juessen (2012) does not feature extreme value shocks, so the moving costs are not exactly a measure of information. Nonetheless, Bayer and Juessen (2012) does estimate substantially smaller moving costs

comparisons tell the reader how much information the model has. The larger the average moving cost, the less the model is able to predict where people will be in the next period, relative to the information of staying in place.[26]

Observation 3 tells us that if the modeler's information set is richer, moving costs will be lower because the modeler can better predict future locations. In Table 2, I investigate whether this predicted relationship holds across papers.[27] Table 2 is roughly ordered by the size of the moving costs, from largest to smallest. The main takeaway from this table is that these moving costs are indeed predictive of the information richness of the model: the lower the moving costs, the more the modeler knows about the potential migrants. In column (4), where the modeler's information is listed, the amount of things that the modeler knows increases as the moving cost decreases.

We can also compare estimated moving costs within the same paper. Zerecero (2021) estimates a model that includes a bias for living in one's birthplace and finds that it features smaller moving costs than a model that does not. This reflects an increase in the information the modeler has available to predict future locations. The Shannon entropy of future locations is smaller when the modeler already knows the person's birthplace. While it is a less direct comparison, Giannone et al. (2023) compares their estimated migration costs to the migration costs in Kennan and Walker (2011) and argues their new costs are lower because they include wealth in their model. This claim is consistent with wealth being an important piece of information about future locations.

Other models also reduce the estimated moving cost by including features that help predict migration. For example, Heise and Porzio (2022) considers job search, where migration is more likely to occur conditional on a job offer, and Porcher (2020) considers rational inattention. Through the lens

($34,248), likely because they incorporate information about migrants persistent preferences over locations.

[26]Having more information does not always make a model "better," as the appropriateness of having information will depend on the economic question.

[27]Of course, the geographies, time periods, and settings are changing as well, so that will also affect the moving costs. As long as these are not correlated to the information set of the modeler, we would still expect a correlation between the information set and estimated moving costs.

of my interpretation, prior to the decision to move, the modeler learns some information—either the agent gets a job offer (Heise and Porzio, 2022) or they pick their optimal signal (Porcher, 2020). From the perspective of the modeler, this information helps predict the agents' future locations, lowering the Shannon entropy. Consistent with my interpretation, the estimated moving costs in these models are lower. Porcher (2020) estimates a model without his information frictions and finds the migration costs are 40 percent higher.

In sum, it does appear that reported moving costs are predictive of the richness of the model, despite the differences in time periods and geography that are also present. So the interpretation of moving costs as a measure of information does help reconcile the different estimates of moving costs in the literature.

4 Conclusion

Many people think of moving costs as a black box, since it is a stand-in for many things that a modeler might not observe: information frictions, job and housing search, psychological costs of relocating, and, of course, the actual monetary costs of moving. I provide an alternative but related interpretation: average moving costs measure the size of a black box; moving costs are related to how little information is in the model about future locations.

Paper	(1) Estimated Migration Costs	(2) Length of time	(3) Geography	(4) Modeler's Information	(5) Notes
Tombe and Zhu (2019)	282% of lifetime income	lifetime	Chinese provinces × urban/rural	birthplace	Paper reports the parameter which I called $\bar{\delta}$ as 2.82 which I interpreted as a share of lifetime income because in their model, moving costs are paid every year a migrant is away from their birthplace
Zerecero (2021)	56% of lifetime consumption	1 year	French départments	current location and birthplace	Without home bias, migration costs estimated to be 10% larger
Bryan and Morten (2019)	39% of lifetime income	lifetime	Indonesian regencies	birthplace	
Bryan and Morten (2019)	15% of lifetime income	lifetime	U.S. States	birthplace	
Ransom (2022)	$394,000 to $459,000 (2004-2013 dollars)	1 year	35 U.S. core-based statistical areas	current location, work experience, age, employment and labor force status, and unobserved type	
Kennan and Walker (2011)	$312,146 (2010 dollars)	1 year	U.S. States	current location, birthplace, current wage, age, type (stayer or mover), last year's location, and wage at that location	
Giannone et al. (2023)	196,202 CAD (2016 dollars)	2 years	Canadian provinces	current location, wealth, income shock, age, housing tenure status, and housing consumption	
Porcher (2020)	75% of annual earnings	1 year	Brazilian mesoregions	current location, information acquired by the agent about productivity in different locations	Without information frictions, migration costs estimated to be 40% larger
Heise and Porzio (2022)	3.1%-5.3% of lifetime income	continuous	4 German regions	current location, home location, current employment status, current wage, location of job offer, wage of job offer	While the model is continuous time, workers only consider moving at discrete times when they get a job offer

Table 2: Migration Costs in the Literature

References

Anas, Alex, "Discrete choice theory, information theory and the multinomial logit and gravity models," *Transportation Research Part B: Methodological*, 1983, *17* (1), 13–23.

Bartik, Alexander W and Kevin Rinz, "Moving costs and worker adjustment to changes in labor demand: Evidence from longitudinal census data," 2018.

Bayer, Christian and Falko Juessen, "On the dynamics of interstate migration: Migration costs and self-selection," *Review of Economic Dynamics*, 2012, *15* (3), 377–401.

Bertoli, Simone, Jesús Fernández-Huertas Moraga, and Lucas Guichard, "Rational inattention and migration decisions," *Journal of International Economics*, 2020, *126*, 103364.

Bishop, Kelly C, "A dynamic model of location choice and hedonic valuation," 2012.

Borusyak, Kirill, Rafael Dix-Carneiro, and Brian Kovak, "Understanding migration responses to local shocks," 2022.

Bryan, Gharad and Melanie Morten, "The aggregate productivity effects of internal migration: Evidence from Indonesia," *Journal of Political Economy*, 2019, *127* (5), 2229–2268.

Caliendo, Lorenzo, Maximiliano Dvorkin, and Fernando Parro, "Trade and labor market dynamics: General equilibrium analysis of the China trade shock," *Econometrica*, 2019, *87* (3), 741–835.

Dix-Carneiro, Rafael, "Trade liberalization and labor market dynamics," *Econometrica*, 2014, *82* (3), 825–885.

Fosgerau, Mogens, Emerson Melo, Andre De Palma, and Matthew Shum, "Discrete choice and rational inattention: A general equivalence result," *International Economic Review*, 2020, *61* (4), 1569–1589.

Giannone, Elisa, Qi Li, Nuno Paixao, and Xinle Pang, "Unpacking moving: A Spatial Equilibrium Model with Wealth," 2023.

Head, Keith and John Ries, "Increasing returns versus national product differentiation as an explanation for the pattern of US–Canada trade," *American Economic Review*, 2001, *91* (4), 858–876.

Heise, Sebastian and Tommaso Porzio, "Labor Misallocation Across Firms and Borders," 2022.

Howard, Greg and Hansen Shao, "The Dynamics of Internal Migration: A New Fact and its Implications," 2025.

Jia, Ning, Raven Molloy, Christopher Smith, and Abigail Wozniak, "The economics of internal migration: Advances and policy questions," *Journal of Economic Literature*, 2023, *61* (1), 144–180.

Kennan, John and James R Walker, "The effect of expected income on individual migration decisions," *Econometrica*, 2011, *79* (1), 211–251.

Kleinman, Benny, Ernest Liu, and Stephen J Redding, "Dynamic spatial general equilibrium," *Econometrica*, 2023, *91* (2), 385–424.

Koşar, Gizem, Tyler Ransom, and Wilbert van der Klaauw, "Understanding Migration Aversion Using Elicited Counterfactual Choice Probabilities," *Journal of Econometrics*, 2022, *231* (1), 123–147. Annals Issue: Subjective Expectations & Probabilities in Economics.

Matějka, Filip and Alisdair McKay, "Rational inattention to discrete choices: A new foundation for the multinomial logit model," *American Economic Review*, 2015, *105* (1), 272–298.

Monras, Joan, "Economic shocks and internal migration," 2020.

Oswald, Florian, "The effect of homeownership on the option value of regional migration," *Quantitative Economics*, 2019, *10* (4), 1453–1493.

Porcher, Charly, "Migration with costly information," 2020.

Ransom, Tyler, "Labor market frictions and moving costs of the employed and unemployed," *Journal of Human Resources*, 2022, *57* (S), S137–S166.

Ruggles, Steven, Sarah Flood, Matthew Sobek, Danika Brockman, Grace Cooper, Stephanie Richards, and Megan Schouweiler, "IPUMS USA: Version 13.0 [dataset]," Online 2023.

Schmutz, Benoît and Modibo Sidibé, "Frictional labour mobility," *The Review of Economic Studies*, 2019, *86* (4), 1779–1826.

Schubert, Gregor, "House price contagion and U.S. city migration networks," 2021.

Shannon, Claude E., "A Mathematical Theory of Communication," *The Bell System Technical Journal*, 1948, *27* (3), 379–423.

Tombe, Trevor and Xiaodong Zhu, "Trade, migration, and productivity: A quantitative analysis of china," *American Economic Review*, 2019, *109* (5), 1843–1872.

Wang, Yao, "Cultural Distance and Internal Migration: Evidence from Indonesia," 2024.

Wilson, Alan Geoffrey, "The use of entropy maximising models, in the theory of trip distribution, mode split and route split," *Journal of Transport Economics and Policy*, 1969, pp. 108–126.

Zerecero, Miguel, "The Birthplace Premium," 2021.

A Proofs

A.1 Proof of Proposition 1

Proof: Plugging in equation (1) to the definition of $\bar{\delta}$ (equation 2),

$$\bar{\delta} = \frac{\sum_{s,i,j:i\neq j} p_i(s) m_{i\to j}(s)(\beta v_{jt}(i,s) - \beta v_{it}(i,s) - \frac{1}{\mu}\log m_{i\to j,t}(s) + \frac{1}{\mu}\log m_{i\to i,t}(s))}{\sum_{s,i,j:i\neq j} p_i(s) m_{i\to j}(s)}$$

Split the fraction into two terms:

$$\begin{aligned}\bar{\delta} =& \frac{\sum_{s,i,j:i\neq j} p_i(s) m_{i\to j}(s)(\beta v_{jt}(i,s) - \beta v_{it}(i,s))}{\sum_{s,i,j:i\neq j} p_i(s) m_{i\to j}(s)} \\ &+ \frac{\sum_{s,i,j,i\neq j} p_i(s) m_{i\to j}(s)(-\frac{1}{\mu}\log m_{i\to j,t}(s) + \frac{1}{\mu}\log m_{i\to i,t}(s))}{\sum_{s,i,j:i\neq j} p_i(s) m_{i\to j}(s)}\end{aligned}$$

The first term is the returns to migration term, $\mathbb{E}^m[\beta v_{jt}(i,s) - \beta v_{it}(i,s)]$. In the denominator of the second term, we substitute in the following definition: $m_i(s) = \sum_{j\neq i} m_{i\to j}(s)$,

$$\bar{\delta} = \mathbb{E}^m[\beta v_{jt}(i,s) - \beta v_{it}(i,s)] + \frac{\sum_{s,i,j,i\neq j} p_i(s) m_{i\to j}(s)(-\frac{1}{\mu}\log m_{i\to j,t}(s) + \frac{1}{\mu}\log m_{i\to i,t}(s))}{\sum_{s,i} p_i(s) m_i(s)}$$

Rearranging,

$$\begin{aligned}\bar{\delta} = \mathbb{E}^m[\beta v_{jt}(i,s) - \beta v_{it}(i,s)] - \frac{1}{\mu}\frac{1}{\sum_{s,i} p_i(s) m_i(s)} \sum_{s,i,j,i\neq j} p_i(s) m_{i\to j}(s) \log m_{i\to j,t}(s) \\ + \frac{1}{\mu}\frac{1}{\sum_{s,i} p_i(s) m_i(s)} \sum_{s,i,j,i\neq j} p_i(s) m_{i\to j}(s) \log m_{i\to i,t}(s)\end{aligned} \tag{10}$$

We will add and subtract $\sum_{s,i} \frac{1}{\mu} \frac{1}{\sum_{s,i} p_i(s) m_i(s)} m_{i\to i}(s) \log m_{i\to i}(s)$ to the right-hand side. We subtract from the middle term and add to the final term. While the equation looks similar, note that the $i \neq j$ condition in the summation is removed.

$$\bar{\delta} = \mathbb{E}^m[\beta v_{jt}(i,s) - \beta v_{it}(i,s)] - \frac{1}{\mu} \frac{1}{\sum_{s,i} p_i(s) m_i(s)} \sum_{s,i,j} p_i(s) m_{i\to j}(s) \log m_{i\to j,t}(s)$$
$$+ \frac{1}{\mu} \frac{1}{\sum_{s,i} p_i(s) m_i(s)} \sum_{s,i,j} p_i(s) m_{i\to j}(s) \log m_{i\to i,t}(s)$$

The middle term can be simplified using the definition of $H(J|i,s) = -\sum_j m_{i\to j}(s) \log m_{i\to j}(s)$. The last term simplifies because $\sum_j m_{i\to j}(s) = 1$.

$$\bar{\delta} = \mathbb{E}^m[\beta v_{jt}(i,s) - \beta v_{it}(i,s)] + \frac{1}{\mu} \frac{1}{\sum_{s,i} p_i(s) m_i(s)} \sum_{s,i} p_i(s) H(J|i,s)$$
$$+ \frac{1}{\mu} \frac{1}{\sum_{s,i} p_i(s) m_i(s)} \sum_{s,i} p_i(s) \log m_{i\to i,t}(s)$$

Recall the definition that $I(i|i,s) = -\log m_{i\to i}(s)$. And note that $\mathbb{E}_{is} m_i(s) = \sum_{s,i} p_i(s) m_i(s)$. Making these substitutions,

$$\bar{\delta} = \mathbb{E}^m[\beta v_{jt}(i,s) - \beta v_{it}(i,s)] + \frac{1}{\mu} \frac{1}{\mathbb{E}_{is} m_i(s)} \sum_{s,i} p_i(s) H(J|i,s) - \frac{1}{\mu} \frac{1}{\mathbb{E}_{is} m_i(s)} \sum_{s,i} p_i(s) I(i|i,s)$$

Reordering the terms, and using the definition of $\mathbb{E}_{is}$,

$$\bar{\delta} = \frac{1}{\mu} \frac{1}{\mathbb{E}_{is} m_i(s)} \mathbb{E}_{is}[H(J|i,s) - I(i|i,s)] + \mathbb{E}^m[\beta v_{jt}(i,s) - \beta v_{it}(i,s)]$$

This is the first formula in the proposition.

To see the alternative way of writing the expression, consider equation (10). Define $m^*_{i\to j}(s) = \frac{m_{i\to j}(s)}{m_i(s)}$ to be the probability of moving to j, conditional on moving at all. Plug in $m^*_{i\to j}(s) m_i(s)$ for $m_{i\to j}(s)$:

$$\begin{aligned}\bar{\delta} =& \mathbb{E}^m[\beta v_{jt}(i,s) - \beta v_{it}(i,s)] \\ &- \frac{1}{\mu}\frac{1}{\sum_{s,i} p_i(s)m_i(s)} \sum_{s,i,j,i\neq j} p_i(s)m_i(s)m^*_{i\to j}(s)(\log m_i(s) + \log m^*_{i\to j,t}(s)) \\ &+ \frac{1}{\mu}\frac{1}{\sum_{s,i} p_i(s)m_i(s)} \sum_{s,i,j,i\neq j} p_i(s)m_{i\to j}(s)\log m_{i\to i,t}(s)\end{aligned}$$

Note that $\sum_{j,j\neq i} m^*_{i\to j}(s) = 1$ and $\sum_{j,j\neq i} m_{i\to j}(s) = m_i(s)$:

$$\begin{aligned}\bar{\delta} =& \mathbb{E}^m[\beta v_{jt}(i,s) - \beta v_{it}(i,s)] \\ &- \frac{1}{\mu}\frac{1}{\sum_{s,i} p_i(s)m_i(s)} \sum_{s,i} p_i(s)m_i(s)\log m_i(s) \\ &- \frac{1}{\mu}\frac{1}{\sum_{s,i} p_i(s)m_i(s)} \sum_{s,i,j,i\neq j} p_i(s)m_i(s)m^*_{i\to j}(s)\log m^*_{i\to j,t}(s) \\ &+ \frac{1}{\mu}\frac{1}{\sum_{s,i} p_i(s)m_i(s)} \sum_{s,i} p_i(s)m_i(s)\log m_{i\to i,t}(s)\end{aligned}$$

Given the definitions of I, H, and $\mathbb{E}^m$,

$$\bar{\delta} = \frac{1}{\mu}\mathbb{E}^m[H(J|i\to \not{i}, s) + I(\not{i}|i,s) - I(i|i,s)] + \mathbb{E}^m[\beta v_{jt}(i,s) - \beta v_{it}(i,s)]$$

□

A.2 Proof of Corollary 1

Proof: Because s is fixed, s' does not depend on the choice of i, so we can write the returns to migration term as $\mathbb{E}^m[\beta v_{jt}(s) - \beta v_{it}(s)]$. Writing it out,

$$\frac{\sum_{i,s}\sum_{j\neq i} p_i(s)m_{i\to j}(s)(\beta v_{jt}(s) - \beta v_{it}(s))}{\sum_{i,s}\sum_{j\neq i} p_i(s)m_{i\to j}(s)}$$

The numerator can be rearranged:

$$\frac{\sum_{i,s}\sum_{j\neq i}(p_i(s)m_{i\to j}(s) - p_j(s)m_{j\to i}(s))\beta v_{it}}{\sum_{i,s}\sum_{j\neq i}p_i(s)m_{i\to j}(s)}$$

Because $\sum_{j\neq i}p_i(s)m_{i\to j}(s) = \sum_{j\neq i}p_j(s)m_{j\to i}(s)$ for all i and s in steady-state, the numerator is zero. This means the average migration rate is equal to only the information term. □

A.3 Proof of Corollary 2

Proof: Under the independence assumption, the returns to migration term can be written as $\mathbb{E}^m[\beta v_{jt} - \beta v_{it}]$. Writing it out,

$$\frac{\sum_{i,s}\sum_{j\neq i}p_i(s)m_{i\to j}(s)(\beta v_{jt} - \beta v_{it})}{\sum_{i,s}\sum_{j\neq i}p_i(s)m_{i\to j}(s)}$$

The numerator can be rearranged:

$$\frac{\sum_{i,s}\sum_{j\neq i}(p_i(s)m_{i\to j}(s) - p_j(s)m_{j\to i}(s))\beta v_{it}(s))}{\sum_{i,s}\sum_{j\neq i}p_i m_{i\to j}}$$

Because $\sum_{s,j\neq i}p_i(s)m_{i\to j}(s) = \sum_{s,j\neq i}p_j(s)m_{j\to i}(s)$ for all i in steady-state, the numerator is zero. This means the average migration rate is equal to only the information term. □

A.4 Proof of Proposition 2

Proof: From equation (1),

$$\delta_{ij}(s) = \beta v_{jt}(i,s) - \beta v_{it}(i,s) - \frac{1}{\mu}\log m_{i\to j,t}(s) + \frac{1}{\mu}\log m_{i\to i,t}(s)$$

If we multiply by $m_{i\to j}(s)$ and sum over i,

$$\sum_i m_{i\to j}(s)\delta_{ij}(s) = \sum_i m_{i\to j}(s)\left(\beta v_{jt}(i,s) - \beta v_{it}(i,s) - \frac{1}{\mu}\log m_{i\to j,t}(s) + \frac{1}{\mu}\log m_{i\to i,t}(s)\right)$$

Similarly, if we consider δ_{ji} and do the same operation,

$$\sum_i m_{i\to j}(s)\delta_{ji}(s) = \sum_i m_{i\to j}(s)\left(\beta v_{it}(j,s) - \beta v_{jt}(j,s) - \frac{1}{\mu}\log m_{j\to i,t}(s) + \frac{1}{\mu}\log m_{j\to j,t}(s)\right)$$

By the assumption in the proposition, the left-hand sides of the above two equations are equal, so the right-hand sides must be equal as well

$$\begin{aligned}\sum_i & m_{i\to j}(s)\left(\beta v_{jt}(i,s) - \beta v_{it}(i,s) - \frac{1}{\mu}\log m_{i\to j,t}(s) + \frac{1}{\mu}\log m_{i\to i,t}(s)\right)\\ &= \sum_i m_{i\to j}(s)\left(\beta v_{it}(j,s) - \beta v_{jt}(j,s) - \frac{1}{\mu}\log m_{j\to i,t}(s) + \frac{1}{\mu}\log m_{j\to j,t}(s)\right)\end{aligned}$$

Multiply by $p_i(s)$ and sum over j and s,

$$\begin{aligned}\sum_{i,j,s} & p_i(s) m_{i\to j}(s)\big(\beta v_{jt}(i,s) - \beta v_{it}(i,s) - (\beta v_{it}(j,s) - \beta v_{jt}(j,s))\big)\\ &= \sum_{i,j,s} p_i(s) m_{i\to j}(s)\Bigg(-\frac{1}{\mu}\log m_{j\to i,t}(s) + \frac{1}{\mu}\log m_{j\to j,t}(s)\\ &\qquad - \left(-\frac{1}{\mu}\log m_{i\to j,t}(s) + \frac{1}{\mu}\log m_{i\to i,t}(s)\right)\Bigg)\end{aligned}$$

Under our assumption that future state variables are independent of the current location,

$$2\sum_{i,j,s} p_i(s) m_{i\to j}(s)\big(\beta v_{jt}(s) - \beta v_{it}(s)\big) = \frac{1}{\mu}\sum_{i,j,s} p_i(s) m_{i\to j}(s)\log\left(\frac{m_{j\to j}(s)}{m_{j\to i(s)}}\frac{m_{i\to j}(s)}{m_{i\to i}(s)}\right)$$

Note that if $i = j$, the terms in the summations on either side are zero. And if we divide by total migration,

$$\frac{2\sum_{i,j,s,i\neq j} p_i(s) m_{i\to j}(s)\big(\beta v_{jt}(s) - \beta v_{it}(s)\big)}{\sum_{i,j,s,i\neq j} p_i(s) m_{i\to j}(s)} = \frac{1}{\mu}\frac{\sum_{i,j,s,i\neq j} p_i(s) m_{i\to j}(s) \log\left(\frac{m_{j\to j}(s)}{m_{j\to i(s)}}\frac{m_{i\to j}(s)}{m_{i\to i}(s)}\right)}{\sum_{i,j,s,i\neq j} p_i(s) m_{i\to j}(s)}$$

Using the definition of $\mathbb{E}^m$,

$$\mathbb{E}^m[\beta v_{jt}(s) - \beta v_{it}(s)] = \frac{1}{2\mu}\mathbb{E}^m\left[\log\left(\frac{m_{j\to j}(s)}{m_{j\to i(s)}}\frac{m_{i\to j}(s)}{m_{i\to i}(s)}\right)\right]$$

□

B Monras (2020) extension

Consider the Monras (2020) model, which features a nested logit formulation, so that the elasticity to move at all is different than the elasticity of where to move to. I will denote the elasticity to move at all with λ and the elasticity of where to move with μ. To keep the notation simple, I will assume there are no additional state variables beyond i. The migration probabilities in his model are given as:[28]

$$\log m_{i\to j} = \mu(\beta v_j - \delta_{ij}) - \mu\beta v_{im} + \lambda\beta v_{im} - \log\left(\exp(\lambda\beta v_i) + \exp(\lambda\beta v_{im})\right) \tag{11}$$

$$\log m_{i\to i} = \lambda\beta v_i - \log\left(\exp(\lambda\beta v_i) + \exp(\lambda\beta v_{im}))\right) \tag{12}$$

$$\log(1 - m_{i\to i}) = \lambda\beta v_{im} - \log\left(\exp(\lambda\beta v_i) + \exp(\lambda\beta v_{im}))\right) \tag{13}$$

where $\beta v_{im} = \frac{1}{\mu}\log\sum_{k\neq i}\exp(\mu(\beta v_k - \delta_{ik}))$. The first step is to solve for $\bar{\delta}$ in terms of observed migration, as in the main text. Subtracting (12) from (11) gives:

$$\log m_{i\to j} - \log m_{i\to i} = \mu\beta v_j - \mu\delta_{ij} - \lambda v_i + (\lambda - \mu)\beta v_{im} \tag{14}$$

Subtracting (12) from (13) gives:

$$\beta v_{im} = \beta v_i + \frac{1}{\lambda}\log(1 - m_{i\to i}) - \frac{1}{\lambda}\log m_{i\to i} \tag{15}$$

Plugging in (15) to (14) gives:

$$\log m_{i\to j} - \log m_{i\to i} = \mu\beta v_j - \mu\delta_{ij} - \mu\beta v_i + (\lambda - \mu)\left(\frac{1}{\lambda}\log(1 - m_{i\to i}) - \frac{1}{\lambda}\log m_{i\to i}\right)$$

Solving for δ_{ij},

$$\delta_{ij} = \beta v_j - \beta v_i - \frac{1}{\mu}\log m_{i\to j} + \left(\frac{1}{\mu} - \frac{1}{\lambda}\right)\log(1 - m_{i\to i}) + \frac{1}{\lambda}\log m_{i\to i}$$

[28] In the main text, Monras (2020) does not include moving costs to simplify the algebra, but they are straightforward to include as I do here.

So the average moving cost is

$$\bar{\delta} = \frac{1}{\mathbb{E}_i m_i} \sum_{i \neq j} p_i m_{i\to j} \left(\beta v_j - \beta v_i - \frac{1}{\mu} \log m_{i\to j} + \left(\frac{1}{\mu} - \frac{1}{\lambda} \right) \log(1 - m_{i\to i}) + \frac{1}{\lambda} \log m_{i\to i} \right)$$

Adding and subtracting $\frac{1}{\mu} \log m_{i\to i}$ and collecting the v's:

$$\begin{aligned} \bar{\delta} =& \frac{1}{\mathbb{E}_i m_i} \sum_{i \neq j} p_i m_{i\to j} \left(-\frac{1}{\mu} \log m_{i\to j} + \frac{1}{\mu} \log m_{i\to i} + \left(\frac{1}{\mu} - \frac{1}{\lambda} \right) (\log(1 - m_{i\to i}) - \log m_{i\to i}) \right) \\ &+ \beta \mathbb{E}^m [v_j - v_i] \end{aligned}$$

The first two terms inside the parentheses are identical to the standard model. So we can plug in the result from the Proposition 1:

$$\bar{\delta} = \frac{1}{\mu} \mathbb{E}^m \left[H(J|i \to \not{i}) + (I(\not{i}|i) - I(i|i)) \right] - \mathbb{E}^m \left[\left(\frac{1}{\mu} - \frac{1}{\lambda} \right) (I(\not{i}|i) - I(i|i)) \right] + \beta \mathbb{E}^m [v_j - v_i]$$

Which simplifies to

$$\bar{\delta} = \mathbb{E}^m \left[\frac{1}{\mu} H(J|i \to \not{i}) + \frac{1}{\lambda} (I(\not{i}|i) - I(i|i)) \right] + \beta \mathbb{E}^m [v_j - v_i]$$

This has a very similar formulation to the second formulation in Proposition 1. But instead of multiplying the Shannon information terms by $\frac{1}{\mu}$, they are multiplied by $\frac{1}{\lambda}$. Intuitively, this makes sense because μ is the elasticity conditional on migrating, and $H(J|i \to \not{i})$ is the Shannon entropy conditional on migrating. Similarly, λ is the elasticity of moving at all and $I(\not{i}|i) - I(i|i)$ is the relative Shannon information of moving to not moving.