Moving Cost Magnitudes in Moving Cost Models

Greg Howard*

February 2, 2024

Abstract

The internal migration literature typically estimates average moving costs to be several times larger than annual income. How should economists interpret this estimate? I show that in the steady-state of a standard model, average moving costs are proportional to the Shannon entropy of next period's location minus the Shannon information of staying in the same location. Therefore, average moving costs are a helpful statistic about the model's predictive power regarding future moves but are not an estimate of the literal cost of moving. This alternative interpretation makes sense of the range of moving costs in the literature.

JEL Codes: R23, D80, J61

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

Many papers in economics estimate the average moving cost to be large, often several times annual household income (Kennan and Walker, 2011; Bryan and Morten, 2019, etc.).[1] This may seem implausibly large when compared to actual expenses associated with a move. 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).[2]

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."[3]

In this paper, I propose a different way to think about these estimated moving costs. I show that average moving costs are a measure of the model's predictive power for agents' future locations. A corollary to this interpretation is that estimated moving costs depend on arbitrary decisions of the modeler, such as the length of a time period or the geographic partition. This corollary rules out a literal interpretation of moving costs and makes the debate about the size of these costs somewhat moot.

To establish my interpretation, I back out moving costs from the observed migration patterns in the data, using formulae implied by a standard moving cost model. I then show algebraically that in steady-state, average moving costs are proportional to the average Shannon entropy of next period's location minus the Shannon information of next period's location being the same

[1] In Table 2, I show a range of large moving cost estimates in the literature.

[2] The literature that uses moving cost models is much larger than the papers that report average moving costs as a main outcome. 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.

[3] Other methodologies of uncovering migration costs also give various different 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.

as the current location (Shannon, 1948). 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. So my result is that moving costs measure how surprised the modeler will be when they find out where an individual lives next year relative to their surprise if they find out that individual did not move.[4]

Based on this result, I show that estimated moving costs change depending on the time period or the geographic partition of the model. A more novel result is that the 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. Because these modeling choices are arbitrary, I argue that interpreting moving costs literally is a mistake.

However, that does not mean moving costs are uninteresting. Given the formulae I provide, one can interpret moving costs as a measure of the information of the model. So comparing moving costs across models is informative of how good those models are at predicting future locations. This alternative interpretation makes sense of some recent results, specifically that richer models of moving—which typically incorporate more information—exhibit smaller moving costs (Zerecero, 2021; Giannone et al., 2023; Heise and Porzio, 2022; Porcher, 2020; Schmutz and Sidibé, 2019).

Of course, if the standard moving cost model were the true model of the world, then moving costs could be interpreted both in the way I describe here and also as a literal average moving cost. Therefore, my argument that migration costs should not be taken literally implies that the standard moving cost model is misspecified. To conclude the paper, I focus on the preference shocks in the model and argue that these are the source of the misspecification. Changing the way they are specified may allow future work to better estimate

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

average moving costs.

Besides the literature that uses moving costs in their models, which I discuss in Section 3, this paper has similarities to the literature that relates discrete choice models to generalized entropy (e.g. Jose, Nau and Winkler, 2008; Fosgerau and de Palma, 2016). These papers show an equivalence between utility maximization and entropy minimization in discrete choice models. To my knowledge, no one has related the estimated moving costs to entropy as I do here.[5] The key assumption that allows me to reach my interpretation of moving costs is that I focus on a setting with a steady-state, where differences in the baseline utilities of locations cancel out.[6]

1 Standard moving cost model

In this section, I use the standard moving cost model to derive an interpretable expression for average steady-state 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.[7]

There is a continuum of people indexed by n that live in discrete locations indexed by i. Time is also discrete and is indexed by t. The population of

[5] Porcher (2020) and Bertoli, Moraga and Guichard (2020) are perhaps the closest papers to this one, in that they have to do with both Shannon entropy and migration. Those papers assume rationally 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. This is equivalent to a discrete choice problem (Matějka and McKay, 2015). However, there is a huge difference from this paper because 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 look like moving costs.

[6] 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.

[7] 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).

people living in i at time t is denoted by p_{it}. The share of people in i who move from i to j at time t is denoted $m_{i\to j,t}$.[8] 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 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 $\mathbb{E}_i$ to refer to the population-weighted average across locations and $\mathbb{E}^m$ to refer to the migration-weighted average. I will be particularly interested in the average migration cost, which we define to be $\bar{\delta} \equiv \mathbb{E}^m[\delta_{ij}]$.

In this section, I assume agents are homogeneous except for their location. I also only consider average moving costs in the steady-state of the model. I relax both of these assumption in Appendix A.

Agents maximize the present value of utility, represented by this value function:

$$V_{nt}(i) = \max_j \log w_{jt} + a_{jt} - \delta_{ij} + \frac{1}{\mu}\epsilon_{jnt} + \beta \mathbb{E} V_{nt+1}(j)$$

where w_{jt} is the (real) wage, a_{jt} is the amenities in j, δ_{ij} is the moving cost from i to j, and ϵ_{jnt} is a time-person-location i.i.d. extreme value shock. β is the discount factor, and μ is a scale parameter, which governs the elasticity of substitution between places.

Define $v_{jt} \equiv \log w_{jt} + a_{jt} + \beta \mathbb{E} V_{nt+1}(j)$. Then the migration rate is given by

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

Because δ_{ii} is normalized to zero, δ_{ij} is given by the following expression:

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

I focus on the following statistic which is often reported in papers in the literature, the migration-weighted average moving cost in the steady-state of

[8]Based on this notation, $m_{i\to i,t}$ will refer to the non-migration rate in i.

the model:[9]

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

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, which is the next period's location. Lower-case 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 (Shannon, 1948). 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 very 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

[9] This summary statistic is not universally reported in papers that estimate moving costs, but it is fairly common (see Table 2 for example). Typically, the averages are migration-weighted to down-weight pairs of locations with little migration and high moving costs.

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.

Proposition 1. *In the steady-state of the standard moving cost model, 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,*

$$\bar{\delta} = \frac{1}{\mathbb{E}_i m_i} \frac{1}{\mu} \mathbb{E}_i[H(J|i) - I(i|i)] \tag{2}$$

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

$$\bar{\delta} = \frac{1}{1 - \sum_i p_i m_{i\to i}} \sum_{i.j:i\neq j} p_i m_{i\to j} \left(-\frac{1}{\mu} \log m_{i\to j} + \frac{1}{\mu} \log m_{i\to i} \right)$$

Note that in steady-state, the v_{it} and the v_{jt}'s all cancel out because $\sum_k p_k m_{k\to i} = \sum_k p_i m_{i\to k}$. The steady-state assumption is a key assumption that allows my interpretation and is why my results do not extend to more-general discrete choice models. Rearranging,[11]

$$\bar{\delta} = \frac{1}{1 - \sum_i p_i m_{i\to i}} \frac{1}{\mu} \sum_i p_i \left(-\sum_j [m_{i\to j} \log m_{i\to j}] + \log m_{i\to i} \right)$$

Recall that I defined $m_i \equiv \sum_{j:j\neq i} m_{i\to j}$ to be the total outmigration from i. Using the definitions of Shannon entropy and Shannon information,

$$\bar{\delta} = \frac{1}{\mathbb{E}_i m_i} \frac{1}{\mu} \mathbb{E}_i[H(J|i) - I(i|i)]$$

□

Proposition 1 establishes that the estimated average moving cost is a mea-

[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.

[11] To derive this expression from the one above, note that $\sum_{j\neq i} m_{i\to j} \log m_{i\to i} = (1 - m_{i\to i}) \log m_{i\to i}$. The $-m_{i\to i} \log m_{i\to i}$ term is then moved into the other term so that the sum is over all j, and not just $j \neq i$. This leaves the $\log m_{i\to i}$ term outside the summation.

sure of information: it is proportional to the expected information of finding out where an individual i will live next period minus the information of finding out that the individual did not move.

In addition to Proposition 1, I can alternatively relate the average moving cost to the Shannon entropy of future locations *conditional on migrating.*

Define an event $\not{i}$ to be when the random variable J takes on any realization that is not i, i.e. a move. I will use the notation $H(J|i \to \not{i})$ to be the conditional Shannon entropy of next period's location given that the agent moves away from i.

Proposition 2. *In the steady-state of the standard moving cost model, average moving costs are the migration-weighted average Shannon entropy of next period's location conditional on moving plus the Shannon information of moving minus the Shannon information of not moving, all divided by the migration elasticity. In math,*

$$\bar{\delta} = \frac{1}{\mu}\mathbb{E}^m \left[H(J|i \to \not{i}) + I(\not{i}|i) - I(i|i)\right] \tag{3}$$

Proof: Define $m^*_{i\to j} = \frac{m_{i\to j}}{m_i}$ to be the probability of moving to j, conditional on moving at all. Then, we can algebraically rearrange the expression for average moving costs as:

$$\bar{\delta} = \frac{1}{\sum_i p_i m_i}\frac{1}{\mu}\sum_i p_i m_i \left(-\sum_{j\neq i}[m^*_{i\to j}\log m^*_{i\to j}] - \log m_i + \log(1-m_i)\right)$$

So

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

□

This formulation is helpful compared to Proposition 1 because it separates out the Shannon entropy conditional on moving from the information involved in moving or not.[12]

[12] 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

The two key assumptions for both propositions are that the model is in steady-state and that the agents are homogenous but for their initial location. In Appendix A, I show that if either of these assumptions is relaxed, then the expressions for $\bar{\delta}$ are very similar to the equations (2) and (3), but with an additional term that represents the average gain from migration net of moving costs and idiosynchratic shocks.[13]

These propositions have three important implications for interpreting migration costs, which I cover in the following three corollaries.

Corollary 1. *Holding the migration elasticity constant, estimated average moving costs depend on the modeler's choice of length of the time period.*

One way to see this corollary is from the formula in Proposition 2.[14] 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 2, 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.

Corollary 2. *Holding the migration elasticity constant, estimated average moving costs depend on the modeler's choice of geographic partition.*

---

which location to move to. The formula becomes

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

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.

[13] I can quantify this term when just the steady-state assumption is relaxed. Using the same data from Section 2, this additional term is quantitatively negligible, about 0.4 percent the size of the terms in equations (2) or (3). Intuitively, one of the reasons this term is so small is because gross migration is much larger than net migration (Jia et al., 2023), so the data is not too far away from steady-state. See Appendix A for details.

[14] Alternative intuition for this corollary can be seen directly in equation (1). 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.

The Shannon entropy of next period's location depends on how the modeler partitions geography.[15] 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.[16] Mechanically, migration rates also increase in the number of locations. As far as I know, 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.

Consider two silly examples to show that moving costs can be estimated to be very large or very small depending on the partition. We will use Proposition 2 for these examples. 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 (3), 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 is

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

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

[15]Again, equation (1) 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.

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

of the population to "move regions." Conditional on "moving regions," the Shannon entropy is zero. So the estimated average moving cost is

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

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* average cost of moving by a factor of more than 5.

Corollary 3. *Holding the migration elasticity constant, estimated average moving costs depend 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 equation (2) becomes[17]

$$\bar{\delta}_s = \frac{1}{\mathbb{E}_{is}m_{is}}\frac{1}{\mu}\mathbb{E}_{is}\left[H(J|i,s) - I(i|i,s)\right] \tag{4}$$

Shannon entropy is convex, and Shannon information is concave, 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 $\bar{\delta}_s$ is weakly smaller that $\bar{\delta}$. If s provides any information about the next periods' location, then the inequality is strict.

This expression also holds for some characteristics that are not immutable. 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.[18] In this setup, I can still derive formula (4).[19] So in a model with consideration sets, the modeler will estimate lower moving costs than in a model without consideration sets.

[17] See Appendix A.3 for the derivation.

[18] As this example illustrates, the characteristic s does not need to be measured in the data. It can be something the modeler can only see with the model.

[19] For the derivation, see Appendix A.4.

While immutable characteristics and consideration sets lead to equation (4), this is not

In the limit, if the modeler knew characteristics that could perfectly predict everyone's next period location (e.g. s was a sufficient statistic for j), then moving costs would be estimated to be negative infinity. This is because not moving, $I(i|i,s)$, would be an infinitely large surprise for someone they could perfectly predict was moving.

All three corollaries assumed that the migration elasticity was held constant, so it worth discussing whether that is a reasonable assumption when changing the length of a time period, the geography, or the modeler's information set. To estimate this elasticity, a modeler typically uses the following equilibrium relationship. They might use an exogenous change in a location's utility as an instrument to estimate μ:

$$d\log m_{i\to j} - d\log m_{i\to i} = \mu(dv_i - dv_j) \tag{5}$$

Importantly, this relationship does not depend that much on how time periods, geographies, or demographic groups are aggregated. I discuss this in Appendix C.

2 Moving cost calibrations with data

In this section, I illustrate the corollaries from the previous section using real world data. In particular, I estimate the average moving costs using equation (2) with data from the Census and the American Community Survey (ACS) in 2000 (Ruggles, Flood, Sobek, Brockman, Cooper, Richards and Schouweiler, 2023).[20] 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

true of all possible characteristics. In Appendix A, I add a general characteristic, which affects migration costs and location utilities. I show that migration costs are the sum of two terms: one which is the difference between the Shannon entropy of next period's location and the Shannon information of not moving, and one that represents the average gain from moving net of migration costs and idiosyncratic shocks. In these two examples, that second term is zero.

[20]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).

Table 1: Estimated Moving Costs for Different Models

	(1) Shannon Entropy	(2) Migration Rate	(3) Estimated Moving Cost	(4) Cost in $1000's
1 year, states	0.182	0.024	6.692	315
	(0.0017)	(0.0002)	(0.0138)	(0.65)
5 year, states	0.561	0.085	5.585	262
	(0.0005)	(0.0001)	(0.0014)	(0.07)
5 year, states (modeler knows birthplace)	0.512	0.085	4.981	234
	(0.0004)	(0.0001)	(0.0018)	(0.08)
5 year, MIGPUMAs	1.231	0.173	5.983	281
	(0.0007)	(0.0001)	(0.0014)	(0.07)

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. The median household income in 2000 (for people also living in the U.S. in 1995) was $47,000, so column (4) is column (3) times 47. Standard errors, in parentheses, are bootstrapped with 100 replications.

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 MIGPUMAs instead of states.[21]

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.[22] 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.[23]

I then calibrate the average moving costs according to equation (2), assuming $\mu = 1$. 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 $\mu = 1$ not because I believe that but because it is easy for the reader to scale the moving costs by whatever μ they prefer.[24] The comparisons of results are intuitive. In the 1 year calibration, I estimate moving costs of 6.7 log points, or when converted to dollars, \$315,000. This is the same order of magnitude as Kennan and Walker (2011), who estimated moving costs of \$312,000 (p. 232).[25]

In the 5 year calibration, I estimate smaller moving costs: 5.6 log points, or \$262,000.[26] This is because at the 5-year horizon, an individual choosing

[21]MIGPUMA stands for Migration Public Use Microdata Area and is a within-state region with at least 100,000 people.

[22]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.

[23]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.

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

[25]The fact that these are only \$3000 different is mostly a coincidence. Kennan and Walker (2011) is using 2010 dollars, while I use 2000 dollars, and the model in Kennan and Walker (2011) is much richer. They also explicitly model a semi-elasticity of migration because they have linear utility in consumption.

[26]This difference is because of the different time horizons, not the different datasets. I

to move is less suprising 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: 5.0 log points ($234,000). This is consistent with Zerecero (2021).

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 10 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 when I use MIGPUMAs instead of states, to 6.0 log points ($281,000).

This means if the "true" model involved drawing an i.i.d. shock for every PUMA, 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.[27]

3 Discussion

3.1 How should moving costs be interpreted?

Is there one of these moving costs that is more "correct" than the other ones? No. The differences depend on arbitrary choices made by the modeler. Natural choices (MIGPUMAs vs. states, 1-year vs. 5-year, and whether to include information on birthplace) lead to large differences in estimated migration costs, of about half a year of income or more. As I showed in Section 1, if

can estimate the standard one-year model using the data from the one-year migration in the ACS, calculate the five-year migration rates from that model, and then estimate the implied moving costs in a five-year model using the five-year simulated data. I estimate a moving cost of 5.110 log points, which is even lower than the number in Table 1.

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

the modeler makes even less natural assumptions, the migration costs could diverge to infinity or negative infinity.

Because these arbitrary decisions affect the estimated moving cost, it therefore cannot be a reasonable measure of the actual cost of moving.

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.[28] These 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.

Table 2 is roughly ordered by the size of the moving costs, from largest to smallest. In column (4), where the modeler's information is listed, the amount of things that the modeler knows increases as the moving cost decreases. Of course, the geographies, time periods, and settings are changing as well, so that will also affect the moving costs, but the information column seems to be a factor.

[28]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. Similarly, Bayer and Juessen (2012) also 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 ($34,248), likely because they incorporate information about migrants persistent preferences over 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

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 to predict future locations. The Shannon entropy, i.e. the average amount the modeler is surprised by any particular location choice, is smaller when they already know 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 location choices and the likelihood of moving.

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 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 at a cost (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.[29] Porcher (2020) estimates a model without his information frictions and finds the migration costs are 40 percent higher.

[29] In Heise and Porzio (2022), I cannot directly apply the formulae in equations (2) and (3) because these models feature additional state variables for the agents. This means that the $v_{it} - v_{jt}$ term from equation (1) will not cancel out. The correct formulae for when there are state variables can be found in Appendix A, and these formulae include an additional term that represents the average utility gain from migration, net of moving costs and idiosynchratic shocks. I expect this additional term would be positive when migration also coincides with a job offer as in Heise and Porzio (2022). So this would actually lead to higher moving costs if there were no change in the modeler's information. Therefore, the decline in migration costs actually understates the improvement in the modeler's information.

3.2 The role of random utility shocks

One assumption of the standard moving cost model—which is key to my interpretation—is that the random utility shocks are i.i.d. extreme value. Of course, shocks like these are necessary to rationalize how otherwise identical agents make different choices. However, the functional form and especially the i.i.d. nature of the shocks leads to my interpretation of the moving costs as a measure of information. It also rules out taking the moving costs literally because moving costs depend on choices of the modeler about time and space. So to make progress on estimating model-based moving costs that correspond to literal moving costs, it is necessary to modify or relax this assumption.[30]

To better understand the role of the i.i.d. assumption, suppose that the standard moving cost model was true, but that the modeler assumes that people draw shocks at the wrong frequency, e.g. every year rather than every five years. Then the researcher incorrectly assumes that in a five-year period, each agent has five draws in which they could get a high enough shock in order to move, whereas in truth, they only get one draw. When moving is rare, this means that for the same migration costs, the researcher would assume about five times more people move using the one-year model than the five-year model. So to match the data, they would infer higher moving costs for the five year model.[31]

A similar argument applies to subgeographies: if a modeler assumes agents get an independent shock for every MIGPUMA, then they will have to assume higher moving costs than if they assume one shock for the whole state in order to match the same amount of migration.[32]

[30]Not all relaxations of this assumption will solve the problem. In appendix B, I show that shocks that generate a nested logit (as in Monras, 2020) still lead to a similar interpretation as in Proposition 2. However, the correlation of shocks in such a model is limited compared to what I propose in this section, e.g. there is no correlation of shocks across time.

[31]In fact, as mentioned in footnote 26, in a simulation based on estimating moving costs using the 1-year ACS data and simulating the model for five years, if the true model is an i.i.d. draw every year, but the modeler incorrectly assumes a draw every five years, the estimated average moving costs will be 5.110, whereas the "true" average moving costs are 6.692.

[32]This point is acknowledged in Kennan and Walker (2011). However, they ignore it because the choice of whether to consider more preference shocks for populous states or

One way to relax this key assumption would be to explicitly model all the heterogeneity, as in Schmutz and Sidibé (2019), where randomness comes from the wage in a job offer that the worker receives, and can be calibrated from data. This approach is quite demanding of the model, as it has to rationalize all the reasons people make different decisions about where to live.

Another approach is to maintain random utility, but to assume individual preferences are correlated across time and space as in Howard and Shao (2022).[33,34] The previous argument would break down because the maximum of several *correlated* random variables is not that different than the average, when the correlation is sufficiently high. A model with these correlations would estimate moving costs that are not as sensitive to the length of time periods or the size of geographies.

Allowing these correlations could also make the model less dependent on the information set of the modeler. For example, knowing the birthplace of an agent means that the modeler is able to predict that the agent will have a larger preference for living near their birthplace, in all time periods. Effectively, the birthplace is inducing a spatial and temporal correlation in the agents' match-specific preferences. If the correlation structure imposed by the modeler were rich enough, it might not matter if the modeler actually knew the agents' birthplaces or not.

---

lower moving costs does not affect their results on income, which was their main result.

[33]In Howard and Shao (2022), there are no moving costs at all, and yet the model matches the data fairly well. This is suggestive that such an approach might estimate lower moving costs.

[34]The standard moving cost model can be thought of as having a specific correlation structure: within-period and within-geography, the correlation of the shocks is 1; and across-periods and across-geographies, the correlation is 0. When you change the length of the period or the size of the geography, the correlation structure changes a lot.

However, in an alternative model like the SPACE model from Howard and Shao (2022), where the shocks are correlated across time and space, then the discrete geographies and time periods can be thought of as an approximation of a continuous model, and so the choice of geographies and time periods will not change the correlation structure that much, provided it is short- and dense-enough to be a good approximation.

4 Conclusion

Many people think of moving costs as a black box, since it is supposed to be 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. In this paper, I provide an alternative but related interpretation: average moving costs measure the size of a black box; moving costs are closely related to how little information is in the model about future locations.

References

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

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 and André de Palma, "Generalized entropy models," 2016.

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

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," 2022.

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.

Jose, Victor Richmond R, Robert F Nau, and Robert L Winkler, "Scoring rules, generalized entropy, and utility maximization," *Operations research*, 2008, *56* (5), 1146–1157.

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

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. Job Market Paper.

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. Job Market Paper.

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.

Zerecero, Miguel, "The Birthplace Premium," 2021. Job Market Paper.

A Extension to the model

In this appendix, I relax two big assumptions that I made in the main text: first, that agents are homogeneous except for their location; and second, that the model is in steady-state.

I derive a general formula for average moving costs, that is the sum of two terms. The first term is the same as in the main text of the paper. The second term is the average change in the continuation value v for migrants, net of migration costs and idiosyncratic shocks.

I then consider a few special cases that are referenced in the main text. First, I consider the case where I drop the steady-state assumption but maintain the homogeneity assumption. I show that this new term is quantitatively small under reasonable assumptions on the symmetry of moving costs. Second, I consider the case where I maintain the steady-state assumption, but allow for the homogeneity assumption to be relaxed based on permanent characteristics of the agents. Finally, I consider the case with the steady-state assumption, but drop the homogeneity assumption to allow for consideration sets.

A.1 More general setup

Consider an extension to the standard moving cost model, in which agents have a state s that affects their payoffs and moving costs. s is multidimensional, and it is a function of both the previous s, the location choice i, and a random variable X. This is a general setup so that s could include age, the history of past locations, job status and wages, etc.

With the state variable, utility is now represented by this value function:

$$V_{nt}(j,s) = \max_i \log w_{it}(s) + a_{it}(s) - \delta_{ji}(s) + \frac{1}{\mu}\epsilon_{int} + \beta \mathbb{E} V_{nt+1}(i, s'(s,i,X))$$

where $w_{it}(s)$ is the (real) wage, $a_{it}(s)$ is the amenities in i, $\delta_{ji}(s)$ is the moving cost from j to i, and ϵ_{int} is an i.i.d. extreme value shock. μ is a scale parameter, which governs the elasticity of substitution between places.

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

$$m_{j\to i,t}(s) = \frac{\exp(\mu(v_{it}(j,s) - \delta_{ji}(s)))}{\sum_k \exp(\mu(v_{kt}(j,s) - \delta_{jk}(s)))}$$

Again, I normalize $\delta_{ii}(s) = 0$, so the $\delta_{ji}(s)$ is then

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

Consider the migration-weighted average moving cost in the steady-state of the model:

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

Rearranging,[35]

$$\begin{aligned}\bar{\delta}_s = & \frac{1}{1-\sum_i p_i(s) m_{i\to i}(s)} \frac{1}{\mu} \sum_i p_i(s)\left(-\sum_j [m_{i\to j}(s)\log m_{i\to j}(s)] + \log m_{i\to i}(s)\right) \\ & + \mathbb{E}^m_{ijs}(v_{it}(j,s) - v_{jt}(j,s))\end{aligned}$$

Define $m_i \equiv \sum_{j:j\neq i} m_{i\to j}$ to be the total outmigration from i. Then

$$\bar{\delta}_s = \frac{1}{\mathbb{E}_{is} m_i(s)} \frac{1}{\mu}\mathbb{E}_{is}\left[H(J|i,s) - I(i|i,s)\right] + \mathbb{E}^m_s[v_{it}(j,s) - v_{jt}(j,s)] \qquad (6)$$

The first term is the same as before, except now the entropy and the information are both conditional on s. The δ is averaged across all s. The second term is the average gain in utility for migrants, net of moving costs and idiosyncratic shocks. The expectation $\mathbb{E}^m_s$ is the average, weighted by the number of migrants of type s moving from i to j.

[35]Note that $\sum_{j\neq i} m_{i\to j} = 1 - m_{i\to i}$.

In this more general setup, average moving costs are the sum of two components: the first is still a measure of the Shannon entropy from the perspective of the modeler minus the Shannon information of not-moving; and the second is the average gains from migration.

In the main text, the second term drops out because there are not average gains to migration. This is because the continuation value is the same for everyone, conditional on location, and I assumed the model was in steady-state, which meant that the same number of people moved into and out of each location.

A.2 Special Case: Not in steady state

If I drop the steady-state assumption, but assume there are no additional s states, then (6) becomes

$$\bar{\delta} = \frac{1}{\mathbb{E}_i m_i} \frac{1}{\mu} \mathbb{E}_i \left[H(J|i) - I(i|i) \right] + \mathbb{E}^m [v_{it}(j) - v_{jt}(j)] \tag{7}$$

The first term is the same as in the main part of the paper, and the second term is the additional utility gains from the fact that there is net migration to better places. The second term is likely to be positive, and it is small when differences in utility across space are small or when net migration is small.

In fact, I can numerically show that they are small, using the same data that I used in Section 2. I assume average moving costs into and out of every location are equal:

$$\sum m_{i \to j} \delta_{ij} = \sum m_{i \to j} \delta_{ji}$$

With this assumption, I can put a number on these average gains from migration net of moving costs and the idiosynchratic utility.[36] This assumption allows me to set up a system of two equations and two unknowns relating $\sum_{j \neq i} m_{i \to j} (v_{it} - v_{jt})$ and $\sum_{j \neq i} m_{i \to j} \delta_{ij}$, based on equation (1). Solving, the

[36]I cannot assume $\delta_{ij} = \delta_{ji}$ for every i and j because it overidentifies the data. With states, there would be 51×51 migration data points, but only 51 v_i's and $\frac{51 \times 50}{2}$ δ_{ij}'s to identify them with.

average gain from migration is given by:

$$\frac{1}{\mu}\sum_{i,j,i\neq j} p_i m_{i\to j}\log\left(\frac{m_{i\to j}}{m_{j\to i}}\frac{m_{j\to j}}{m_{i\to i}}\right)$$

In the data, and with $\mu = 1$, this number is about 0.022. This is about 0.4 percent of the size of the information term (see Table 1). So at least in the standard model, the steady-state assumption was not quantitatively affecting my results.

A.3 Special Case: s is immutable

Another special case of the more general result is if s is immutable: i.e. $s' = s$, and I maintain the steady-state assumption. s being immutable means I can rewrite $v_{it}(j, s) \equiv v_{its}$, i.e. the continuation value does not depend on j. And because the model is in steady-state, the number of people of type s moving into i is cancelled out by the number of people of type s moving out of i. So the average gains from migration term drops out:

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

This leaves us with the main result again, but where the Shannon entropy and the Shannon information are conditional on s.

A.4 s and j do not affect v_{it}

Another straightforward example is if $v_{it}(j, s)$ does not depend on j or s, e.g. s governs the contemporaneous moving costs, but nothing else.[37] This could be the case if, at the start of each period, each agent drew a random consideration set, which is represented by s. But once they moved to the new region, they looked just like anyone else there.

[37] As in the last example, I maintain the steady-state assumption here.

Under this assumption, we again get the same equation:

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

So the Shannon entropy and Shannon information depend on s, but the result is otherwise the same.

In general, however, adding the state to the model leads to the possibility that there are average gains to in utility for migrants. Examples of s that would matter are if s equals the history of past locations, age, or employment status. I would expect these to be positive because migrants will tend to move to places with higher utility for themselves.

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. The migration probabilities in his model are given as:[38]

$$\log m_{i\to j} = \mu(v_j - \delta_{ij}) - \mu v_{im} + \lambda v_{im} - \log\left(\exp(\lambda v_i) + \exp(\lambda v_{im})\right) \quad (8)$$

$$\log m_{i\to i} = \lambda v_i - \log\left(\exp(\lambda v_i) + \exp(\lambda v_{im}))\right) \quad (9)$$

$$\log(1 - m_{i\to i}) = \lambda v_{im} - \log\left(\exp(\lambda v_i) + \exp(\lambda v_{im}))\right) \quad (10)$$

where $v_{im} = \frac{1}{\mu}\log\sum_{k\neq i}\exp(\mu(v_k - \delta_{ik}))$. The first step is to solve for $\bar{\delta}$ in terms of observed migration, as in the main text. Subtracting (9) from (8) gives:

$$\log m_{i\to j} - \log m_{i\to i} = \mu v_j - \mu\delta_{ij} - \lambda v_i + (\lambda - \mu)v_{im} \quad (11)$$

Subtracting (9) from (10) gives:

$$v_{im} = v_i + \frac{1}{\lambda}\log(1 - m_{i\to i}) - \frac{1}{\lambda}\log m_{i\to i} \quad (12)$$

Plugging in (12) to (11) gives:

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

Solving for δ_{ij},

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

[38] 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

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

In steady-state, the v_i and v_j all cancel out, as in the main text:

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

Adding and subtracting $\frac{1}{\mu} \log m_{i\to i}$,

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

The first two terms inside the parentheses are identical to the standard model. So we can plug in the result from the Proposition 2:

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

Which simplies to

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

This has a very similar formulation to Proposition 2. 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.

C Estimating μ

In this section, we consider whether equation (5) would change very much depending on the aggregation of time periods, geographies, or demographics.

First, let us consider changing the length of a time period. Define $m^t_{i\to j}$ to be the migration rate over time horizon t. In practice, $m^t_{i\to j}$ is a multiplicative function of a baseline migration rate and a function of the time horizon being considered, in both standard models and in the data. In particular, the standard model implies $m^t_{i\to j} \approx m^1_{i\to j} t$ when moving costs are sufficiently large, and in the data, $m^t_{i\to j} \approx m^1_{i\to j}\sqrt{t}$ (Howard and Shao, 2022).[39] Because of the multiplicative property,

$$d\log m^t_{i\to j} \approx d\log m^1_{i\to j}$$

This implies equation (5) is approximately time horizon invariant. Based on equation (5), we should also be cognizant of $d\log m_{i\to i}$, but proportional changes in the non-migration rate are tiny compared to changes in migration rates.

Second, we can consider aggregating to a larger geography. Suppose we combine two locations j and k into ℓ. If previously, equation (5) held for j and k, then we can add them together based on initial migration levels:

$$\begin{aligned}\frac{m_{i\to j}}{m_{i\to j}+m_{i\to k}} d\log m_{i\to j} + \frac{m_{i\to k}}{m_{i\to j}+m_{i\to k}} d\log m_{i\to k} - d\log m_{i\to i} \\ = \mu\left(\frac{m_{i\to j}}{m_{i\to j}+m_{i\to k}} dv_j + \frac{m_{i\to j}}{m_{i\to j}+m_{i\to k}} dv_k - dv_i\right)\end{aligned}$$

The left-hand side is a log-linear approximation of $d\log m_{i\to\ell}$. Assume that

$$dv_\ell \approx \frac{m_{i\to j}}{m_{i\to j}+m_{i\to k}} dv_j + \frac{m_{i\to j}}{m_{i\to j}+m_{i\to k}} dv_k$$

This assumption will hold if the wage increases are similar in j and k or if the

[39]The fact that the data and the standard model do not match is one of the main contributions of Howard and Shao (2022), but is immaterial for this argument as long as some function of t can be pulled out.

population weights and the migration weights are similar. Then the previous expression reduces to

$$d \log m_{i \to \ell} - d \log m_{i \to i} \approx \mu(dv_\ell - dv_i)$$

In other words, we still expect equation (5) to hold approximately.

Finally, we can consider aggregating across demographics. Assume that equation (5) held for two groups s and r, which we then aggregate.

$$\begin{aligned} &\gamma_s(d \log m_{i \to j,s} - d \log m_{i \to i,s}) + \gamma_r(d \log m_{i \to j,r} - d \log m_{i \to i,r}) \\ &\quad = \mu\left(\gamma_s(dv_{j,s} - dv_{i,s}) + \gamma_r(dv_{j,r} - dv_{i,r})\right) \end{aligned}$$

where $\gamma_s = \frac{p_{i,s} m_{i \to j,s}}{p_{i,s} m_{i \to j,s} + p_{i,r} m_{i \to j,r}}$ and $\gamma_r = \frac{p_{i,r} m_{i \to j,r}}{p_{i,s} m_{i \to j,s} + p_{i,r} m_{i \to j,r}}$, i.e. weights corresponding to the amount of migration from each group.

The left-hand side is again a log-linear approximation of the total change in migration across r and s, so as long as

$$dv_\ell \approx \gamma_s dv_{\ell,s} + \gamma_r dv_{\ell,r}$$

for $\ell = i, j$ then the previous expression reduces to

$$d \log m_{i \to j} - d \log m_{i \to i} \approx \mu(dv_j - dv_i)$$

Hence, we should not expect the estimation of μ to vary with geographic partition, time horizon, or demographics.