Moving Cost Magnitudes in Moving Cost Models

Greg Howard*

October 13, 2023

Abstract

What is the correct interpretation of a moving cost? I show that average moving costs, in the steady-state of a standard moving cost model, are proportional to the difference between the Shannon entropy of next period's location and the Shannon information of not moving. Therefore, moving costs are correctly interpreted as a statistic about the modeler's lack of information regarding future moves, but not about the actual cost of moving. This alternative interpretation helps make sense of the wide range of moving costs in the literature.

*University of Illinois, Urbana-Champaign. All errors are my own.

Many papers in economics estimate the average cost of moving to be quite large, often several times annual household income (Kennan and Walker, 2011; Dix-Carneiro, 2014; Zerecero, 2021; Giannone, Li, Paixao and Pang, 2023).[1] This may seem implausibly large when compared to actual expenses associated with a move. However, others have noted that these migration costs could reflect other frictions, too, and that if we explicitly model these other frictions, then estimated moving costs fall (Schmutz and Sidibé, 2019; Heise and Porzio, 2019; Porcher, 2020).[2]

Jia, Molloy, Smith and Wozniak (2023), a review article in the *Journal of Economic Literature*, summarizes it 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."

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 modeler's information about future locations. A corollary to this interpretation is that estimated moving costs depend on arbitrary decisions of the modeler, such as the time period or the geographic partition. This corollary rules out being able to interpret moving costs literally and makes the debate about the size of these costs somewhat moot.

To establish my interpretation, I show algebraically that in the steady-state of a standard moving cost model, moving costs are proportional to the average Shannon entropy of next period's location minus the Shannon information

[1] Kennan and Walker (2011) reports average moving costs for all possible moves, but reports moving costs paid by actual movers that are much lower, in fact negative. This is because the i.i.d. preference shocks that I discuss in the next section are included in their measure of moving costs. However, their most well-known statistic is measured similarly to the statistic I consider below: "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).

[2] Perhaps because of this debate, the literature that uses moving costs in their models is much larger than the number of papers that report moving costs as a main outcome. For example, Caliendo, Dvorkin and Parro (2019) has a similar model that would imply similarly sized moving costs, but develops solution techniques that do not require backing out the moving cost parameters. Schubert (2021) does not report the average moving cost, but does consider counterfactuals in which the moving costs change.

of not moving.[3] In other words, moving costs are a measure of information: specifically, how surprised the modeler is by the agent's future location compared to the agent not moving.

Based on this result, I show that moving costs are sensitive to seemingly arbitrary assumptions that the modeler makes about how to partition geographies or the length of a time period. They are also sensitive to the modeler's information about the people in the model. For example, knowing the birthplace of each person leads the modeler to estimate smaller moving costs. I give examples of the ways these arbitrary decisions affect moving costs using data from the 2000 Census and the American Community Survey. Because none of these modeling choices affects the actual decisions of people on the margin of moving, I argue that interpreting moving costs literally is a mistake.

However, that does not mean moving costs are uninteresting. Given the formula I provide, one can interpret moving costs as a measure of the information of the modeler. 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, e.g. that richer models of moving—which typically incorporate more information—exhibit smaller moving costs (Zerecero, 2021; Giannone et al., 2023; Heise and Porzio, 2019; Porcher, 2020; Schmutz and Sidibé, 2019).

I conclude the paper with a discussion of the assumptions in the model that lead to my interpretation, and the potential for future research that could get at better estimates of actual moving costs.

Besides the literature that uses moving costs in their models, which I discussed throughout the introduction, this paper also has some similarities to the literature that relates discrete choice models to generalized entropy (e.g. Jose, Nau and Winkler, 2008; Fosgerau and de Palma, 2016). To my knowledge, while people have shown many interesting relationships between utility maximization and entropy, no one has related the estimated moving costs to entropy as I do here.[4]

[3]The concepts if Shannon entropy and Shannon information were developed in Shannon (1948).

[4]Porcher (2020) is perhaps the closest paper to this one, in that it has to do with both

1 Standard Moving Cost Model

Consider the standard moving cost model, in which agents are maximizing the present value of utility, represented by this value function:

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

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

If we define $v_{it} \equiv \log w_{it} + a_{it} + \beta \mathbb{E} V_{nt+1}(i)$, then migration is given by

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

Typically δ_{ii} is normalized to zero, allowing us to write the following expression for δ_{ji}:

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

Intuitively, the less migration we see, the higher we infer the moving costs are.

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

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

Note that in steady-state, the v_{it} and the v_{jt}'s all cancel out because $\sum_k p_k m_{k \to i} =$

Shannon entropy and internal migration. That paper assumes rationally inattentive agents, and a typical result in 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 between Porcher (2020) and this paper because this paper emphasizes the moving costs as a measure of the modeler's lack of information, whereas that paper emphasizes that agents' lack of information can look like moving costs.

$\sum_k p_i m_{i\to k}$.[5] Rearranging,[6]

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

If we define $m_i \equiv \sum_{j:j\neq i} m_{i\to j}$ to be the total outmigration from i, then

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

where $H(j|i)$ is the Shannon entropy of a person's location next period, given their location this period, and $I(j = i)$ is the Shannon information of not moving.[7] $\mathbb{E}_i$ refers to the population-weighted average across states. The term within the expectation can be thought of as the difference between the Shannon entropy of the future location and the Shannon information of not moving.

An easy way to understand Shannon information is as a measure of 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. Sometimes, people like to think of Shannon entropy as 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

[5] Of course, when we calibrate the model using the data from any given year, the model is not in steady-state. However, in Appendix A, we show that if the model is not in steady-state, our main equations (2) and (3), are very similar, but with one additional term that represents the average gain from migration net of moving costs and idiosynchratic shocks. When we take it to the data, it is quantitatively negligible, about 0.4 percent the size of the steady-state terms we focus on. Intuitively, one of the reasons this term is so small is because gross migration is much larger than net migration (Jia et al., 2023).

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

[7] For a set of mutually exclusive events $i \in I$, Shannon entropy is defined to be $H(i) = -\sum_{i\in I} \pi_i \log \pi_i$, where π_i is the probability of each possible i. Shannon information of an event with probability π_i is given by $-\log \pi_i$. Note that Shannon entropy is the *ex ante* expectation of Shannon information.

contains.[8]

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

In addition to the previous result of equation (2), we can alternatively relate the average moving cost to the Shannon entropy of future locations *conditional on migrating*. 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, another way to express the same average moving cost is:

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

So

$$\bar{\delta} = \frac{1}{\mu} \mathbb{E}^m_i \left[H(j|i, i \neq j) + I(j \neq i) - I(j = i) \right] \tag{3}$$

where $H(j|i, i \neq j)$ is the Shannon entropy of tomorrow's location, given today's location, and that tomorrow's location is not the same as today's location; $I(j \neq i)$ is the information content of moving anywhere, and $I(j = i)$ is the information content of staying. $\mathbb{E}^m$ signifies that the average for this equation is weighted based on the number of migrants, not the total population as in equation (2).

Therefore, the average moving cost is proportional to the difference in the information of moving, plus the entropy of location conditional on moving, and the information of not moving. Each bit of expected information "costs" $\frac{1}{\mu \log 2} w$ per migrant.

Based on these formulations, I make three observations:

Observation 1. *Average moving costs depend on the modeler's choice of time*

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

horizon.

Over short time horizons, the Shannon entropy, conditional on moving, does not vary much. Whether we look at 1 year or 5 years, the percentage of migrants from state i moving to state j, $m^*_{i\to j}$ is roughly constant. However, migration rates are smaller for shorter time horizons. So based on equation (3), we can see that average moving costs will vary with the time period chosen as the second term, $I(j \neq i) - I(j = i) = \log \frac{1-m_i}{m_i}$, will increase when time horizons are short. In fact, as time horizons get arbitrarily short, estimated average moving costs will get arbitrarily large.[9]

Observation 2. *Average moving costs depend on the modeler's choice of geographic partition.*

The Shannon entropy of next period's location depends on how the modeler partitions geography. 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.[10] 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 our empirical results, we show that the Shannon entropy change dominates the change in the information of not moving when we 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.

In fact, we can consider two silly examples to show that moving costs can be estimated to be very large or very small depending on the partition. In

[9]When we express migration costs in utility terms, one might be worried that the correct way to interpret those costs also depend on the time horizon. For example, when we convert to dollar terms, and we consider five-year time periods, we will calibrate w to five times the yearly wage, but when we consider one-year time periods, we will calibrate w to the yearly wage. However, unless migration over time horizon t is proportional to $\exp(t)$, this will not cancel. Unfortunately, in the data, it is closer to proportional to $\sqrt{t}$ (Howard and Shao, 2022).

[10]This does not have to be true in all cases, if the more precise location is extremely informative of which locations the person is likely to move to.

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. But 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 $\log(57/43)$. 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 on this enormous number would be $-\log\frac{1}{1000} + \log\frac{57}{43} \approx 7.2$.

Alternatively, we could partition the United States into houses with an even-numbered address and ones with an odd-numbered address. If we assume that it is random which type of house you move into, we would expect 21.5 percent of the population to be "moving regions" each year. But conditional on moving, the Shannon entropy is zero. So the estimated average moving cost would be $\log\frac{78.5}{21.5} \approx 1.3$.

Observation 3. *Average moving costs depend on the modeler's information set.*

Suppose we knew some immutable characteristic about people, denoted by s, such as their race or their birthplace. If we estimate separate moving costs by this characteristic, then equation (2) becomes

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

Shannon entropy is convex, and Shannon information is concave, so by Jensen's inequality, $\mathbb{E}_{is}[H(j|i,s) - I(j=i|s)] \leq \mathbb{E}_i H(j|i) - I(j=i)]$. Since $\mathbb{E}_{is} m_{is} = \mathbb{E}_i m_i$, the whole quantity, $\bar{\delta}_s$ is weakly smaller that $\bar{\delta}$. And if s provides any information about the next periods' location, then the inequality will be strict.

In the limit, if we could perfectly predict everyone's next period location (e.g. $s=j$), then moving costs would be reduced to negative infinity. This is because not moving would be an infinitely large surprise for someone you knew was moving.

Next, we consider s's that may not be an immutable characteristic of the person. For example, if the modeler modeled the decision making process in two stages where, first, each person came up with a consideration set, and second, compared the utilities available in each, s could be the consideration set.[11] In this setup, we can still derive formula (4).[12] So in a model with consideration sets, the modeler will estimate lower moving costs than in a model without consideration sets.

2 Moving Cost Calibrations with Data

In this section, I take the observations from the last section and illustrate them using real world data.

In particular, we want to estimate the average moving costs using equation (2) with data, using the Census and the American Community Survey from 2000 (Ruggles, Flood, Sobek, Brockman, Cooper, Richards and Schouweiler, 2023).[13] 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 MIGPUMAs instead of states.[14]

I then calibrate the average moving costs according to equation (2), assuming $\mu = 1$. In the literature, there is little consencus 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

[11] 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.

[12] While immutable characteristics and consideration sets do not break the main results, this is not true of all possible characteristics. In Appendix A, we add a general state variable to each person, which can affect their migration costs and location utilities. In general, we 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 idiosynchratic shocks. In these two examples, that second term is zero, but that does not need to generally be true.

[13] 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).

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

Table 1: Estimated Moving Costs for Different Datasets

	Shannon Entropy	Migration Rate	Estimated Moving Cost	Cost in $1000's
1 year, states	0.182	0.024	6.692	315
5 year, states	0.561	0.085	5.585	1312
5 year, states (modeler knows birthplace)	0.512	0.085	4.981	1171
5 year, MIGPUMAs	1.231	0.173	5.983	1406

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 for 1-year migration, the last column is that times the estimated moving cost. For 5-year migration, the last column is $47,000 times five times the estimated moving cost.

because it is easy for the reader to scale the moving costs by whatever μ they prefer.[15] 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 not too different from Kennan and Walker (2011), who estimated moving costs of $312,000 (p. 232).[16] In the 5 year calibration, I estimate substantially smaller moving costs because many more people are moving: 5.6 log points.[17] However, 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. This is consistent with Zerecero (2021)

[15] 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 typically highly correlated across space, and therefore affect both origin and destination locations for many migrants. They suggest a non-linear least squares estimation instead.

[16] 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 the elasticity of migration, which in their paper is a semi-elasticity since they have linear utility in consumption.

[17] In dollars, this number is actually larger because we multiply by five years of wages.

who finds lower estimated moving costs in a model with birthplace. Finally, if we 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 we use MIGPUMAs instead of states, to 6.0 log points.

3 Interpretation

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. Even very natural choices, such as 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 log-point or more. As we showed in Section 1, if we allow the modeler to make even less "normal" 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? They should not try to gauge if the moving cost is big or small compared to their prior on how expensive it is to move. However, they may want to compare the moving costs to other papers or other model specifications. These comparisons can give a sense of 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. For example, 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 the 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.[18]

Other models also reduce the estimated moving cost by including features that help predict migration. For example, Schmutz and Sidibé (2019) and Heise and Porzio (2019) consider 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 (Schmutz and Sidibé, 2019; Heise and Porzio, 2019) or they pick their optimal signal at a cost (Porcher, 2020). From the perspective of the modeler, this information helps predict where the agents are going to move to, lowering the Shannon entropy. Consistent with my interpretation, the estimated moving costs in these models are much lower.[19]

3.1 Why is the standard model bad for estimating true moving costs?

One of the key assumptions of the standard moving cost model is that the shocks are i.i.d. This assumption is clearly unrealistic because it relies on there being no spatial correlation or persistence in the unobserved reasons that people choose locations. Nonetheless, two of the previous observations depend on this i.i.d. assumption.

To see this direct connection, if the researcher assumes that people draw

[18] Giannone et al. (2023) and Kennan and Walker (2011) also differ in their geographies and timing, so the information effect of wealth is not the only difference between the two papers.

[19] In these models, we cannot directly apply the formulas 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 formulas for when there are state variables can be found in Appendix A, and these formulas 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 Schmutz and Sidibé (2019) and Heise and Porzio (2019). 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.

i.i.d. shocks at the wrong frequency, i.e. every year rather than every five years, then they assume 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 if the truth is every five years, 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.

A similar argument can be made for subgeographies: if a researcher 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 in the data.

However, if the researcher relaxed the i.i.d. assumption and instead assumed the shocks were highly correlated across time and space, then this argument would break down because the maximum of several random variables is not that different than the average, when the correlation is sufficiently high.

Trying to get an estimate of "true" moving costs, that are not sensitive to timing and geographic assumptions, would require a model with both geographic and temporal correlation in location-person-match-specific preferences. This would be an interesting direction for future research.

4 Conclusion

In this paper, I show that the moving costs that are estimated by moving cost models measure the average Shannon entropy of next period's location minus the Shannon information of not moving. This measure is sensitive to seemingly-arbitrary choices of the economic modeler, and so I argue that it is not a good measure of actual moving costs. But what it does measure is nonetheless interesting, and helps us understand some of the various measures of moving costs in the literature.

References

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

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, "Spatial wage gaps and frictional labor markets," *FRB of New York Staff Report*, 2019, (898).

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.

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.

Porcher, Charly, "Migration with costly information," 2020. Job Market Paper.

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.

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

A Extensions to the Model

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:

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

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

If we 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

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

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:

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

Rearranging,[20]

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

If we define $m_i \equiv \sum_{j:j\neq i} m_{i\to j}$ to be the total outmigration from i, then

$$\bar{\delta} = \frac{1}{\mathbb{E}_{is} m_i(s)} \frac{1}{\mu} \mathbb{E}_{is} \left[H(j|i,s) + I(j = i|s) \right] + \mathbb{E}^m_{is}[v_{it}(j,s) - v_{jt}(j,s)] \tag{5}$$

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

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 we assumed we were in steady-state, which meant that the same number of people moved into and out of each location.

If we drop the steady-state assumption, but still assume there are no additional s states, then

$$\bar{\delta} = \frac{1}{\mathbb{E}_i m_i} \frac{1}{\mu} \mathbb{E}_i \left[H(j|i) + I(j = i) \right] + \mathbb{E}^m_{ij}[v_{it}(j) - v_{jt}(j)] \tag{6}$$

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

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

better places. The second term is almost always positive and it is small when differences in utility across space are small or when gross migration is much larger than net migration.

If we assume average moving costs into and out of every location are equal: i.e. $\sum m_{i\to j}\delta_{ij} = \sum m_{i\to j}\delta_{ji}$, we can actually put a number on these average gains from migration net of moving costs and the idiosynchratic utility.[21] This assumption allows us 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 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, which we showed in Table 1. So at least in the standard model, the steady-state assumption was not quantitatively affecting our results.

There are several other scenarios in which the average gain in utility for migrants is zero, and the term drops out, two of which we discuss in the main text.

The first is if s is immutable: i.e. $s' = s$, and we maintain the steady-state assumption. s being immutable means we can rewrite $v_{it}(j,s) \equiv v_{its}$, i.e. the continuation value does not depend on j. And because we are 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:

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

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. This could be the case if, at the start of each period, each agent drew a random

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

consideration set, which is represented by s. But once they moved to the new region, they looked just like anyone else there.

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. We would expect these to be positive because migrants will tend to move to places with higher utility for themselves.