The Dynamics of Internal Migration: A New Fact and its Implications*

Greg Howard† Hansen Shao‡

March 2, 2026

Abstract

We propose a new model of internal migration, based on persistent and spatially-correlated idiosyncratic utility. The model is motivated by a new fact in the data that simple moving cost models struggle to match: the t-year interstate migration rate is proportional to the square root of t. The new model maintains the tractability and flexibility of standard migration models, but better matches the dynamics of migration, including the new fact. It has substantially different welfare implications and makes different counterfactual predictions, especially in terms of dynamic adjustment and long-run responses.

Keywords: regional evolution, misallocation, gravity equation, labor mobility, moving costs

JEL Codes: R23, R13, J61

*This paper was previously circulated as "Internal Migration and the Microfoundations of Gravity." We would like to thank Javier Quintana, Kyle Mooney, Yao Wang, Vivek Bhattacharya, Gregor Schubert, Andrii Parkhomenko, Eduardo Morales, Marieke Kleemans, Kyuri Park, Cihang Wang, Jaysa Rafi, Humberto Martinez-Garcia and participants at the Illinois Macro Lunch, Illinois Young Applied Faculty Lunch, the European UEA meetings, Midwest Macro, the AREUEA National Meetings, the Stanford Institute for Theoretical Economics Housing & Urban Economics Conference, the Online Spatial and Urban Seminar, the Chicago Fed Urban Conference, the Cleveland Federal Reserve, and the North American UEA meetings for constructive feedback. We also would like to thank Jialan Wang, Julia Fonseca, and Peter Han for creating the Gies Consumer and Small Business Credit Panel and the Gies College of Business for supporting this dataset.

†University of Illinois Urbana-Champaign, glhoward@illinois.edu.

‡University of Illinois Urbana-Champaign.

1 Introduction

We document a new empirical regularity: the t-year interstate migration rate, defined as the share of people living in a different state than they did t years ago, scales quite closely with the square root of t. This new fact is a puzzle for the widely used moving cost model, which typically implies a linear relationship.

The main contribution of our paper is a simple but novel model that can match the new fact by assuming that idiosyncratic utility is correlated across time and space. The model we propose features Spatially and Persistently Autocorrelated Epsilons, so we refer to it as the "SPACE" model.[1] Unlike the standard model, which interprets low migration rates as the result of large moving costs, our model rationalizes low migration rates as a result of high persistence in idiosyncratic determinants of location. Our model is able to replicate bilateral one-year migration flows from the data, maintaining the flexibility of the standard models, while also featuring more realistic dynamics.

We compare the implications of the new model to the moving cost model and find that, for many but not all questions, the models draw different conclusions. These findings reshape our understanding of the causes of low migration, the dynamics of population adjustment, the long-run population elasticities to local changes, and the changes in implied utilities across space in recent years.

The first part of our paper documents the new square root fact using data from the Gies Consumer and Small Business Credit Panel (GCCP), a 15-year panel recording the location of approximately 1 percent of all Americans with a credit report every year. While the square root fact is related to the well-known fact that return migration is common (Kennan and Walker, 2011), we show that the $\sqrt{t}$ fact is not a simple result of return migration but captures richer dynamics. We also show that this fact does not naturally occur in standard moving cost models. In those models, location choice is a Markov process, which, combined with low rates of migration, implies that the t-year migration rate has an approximately linear relationship with t.[2]

We then build a model that can reconcile this new fact while maintaining the tractability and flexibility of the standard model. Building on the model of McFadden (1978), we

[1] ϵ is the common notation for the random component in a random utility model.

[2] With flexible enough moving costs that depend on the past history of location choices, a model could match almost any relationship between the t-year migration rate and t. Nonetheless, we still think the square root fact is a puzzle for two reasons: first, the models that people actually use do not explain it, possibly due to the large state-space of very flexible models; and second, because when a model can match any relationship, there is no reason for it to match this particular relationship. Our model provides a rationale for this particular relationship. We discuss this further in Section 2.3.

consider a generalized-extreme-value discrete-choice model to introduce correlation over space and time. The SPACE model leads to closed-form solutions for state populations and interstate migration. One important result of this model is that the cross-state population elasticities—a key statistic for quantitative spatial modeling—are directly proportional to the bilateral migration rate. We show a way to calibrate the model that allows the use of simple formulas for population changes, i.e. "exact hat" algebra, and which also allows computationally feasible simulations of individuals' location choices over time. This tractability allows for the SPACE model to easily serve as the migration block of more complex quantitative dynamic spatial models.[3]

Next, we compare the implications of the SPACE model to those of the moving cost model. For many questions, we show that choosing how to model migration is not innocuous but leads to important differences in how economists answer central questions about location choice and migration.

First, we demonstrate that the SPACE model is better at predicting future locations of individuals. We compare the forecasting performance of each model using the Kullback-Leibler divergence, and demonstrate that the SPACE model does better at predicting out-of-sample locations, especially at longer horizons. This is consistent with the idea that the SPACE model is able to match realistic dynamics of location choice.

Second, the SPACE model and the moving cost model have very different perspectives on why people do not move. Moving cost models estimate large moving costs (e.g. Kennan and Walker, 2011; Bryan and Morten, 2019; Giannone, Li, Paixao and Pang, 2020; Zerecero, 2021). In contrast, the SPACE model does not need moving costs at all to rationalize observed levels of migration in the data.

Third, we turn to macroeconomic questions, which typically depend on the elasticity of local populations to local utility. We show that both models feature similar short-run population cross-elasticities, in that the elasticity of population in state i with respect to utility in state j is approximately proportional to the gross migration rate between the two states.[4] In other words, if the purpose of a model with migration is to predict short-run effects on populations, both models deliver similar results.

[3]For example, in Appendix E, we embed the SPACE model into a model with local housing production to study transition dynamics.

[4]Even though the models have similar elasticities, the rationale for why the elasticity is related to migration is a bit different. In the SPACE model, the rationale is that migrants are close to indifferent between living in each state, so the mass of people who will move in response to a small shock is proportional to the number of migrants. In the moving cost model, the rationale is that the extreme value function has a functional form such that the number of people on the margin is proportional to the number of people who make that choice.

However, in the long run, population cross-elasticities are quite different across the two models. In the SPACE model, population elasticities are the same in the short-run and the long-run, meaning that long-run elasticities are still proportional to the migration between the two places. But in the moving cost model, the long-run population elasticities are approximately the same as a static discrete choice logit model, i.e. the elasticity is proportional to the population share of the shocked region. In the data, population shares and gross migration rates have little correlation, so the long-run elasticities of the two models are also uncorrelated.

A fourth difference follows naturally from the previous one: the dynamics of regions' population changes are quite different in the two models. In the SPACE model, the dynamics are simple. In response to a permanent utility change, the population adjusts fully, contemporaneous to the utility shock. But in the moving cost model, the dynamics are relatively slow and can be unintuitive. In that model, every period, a new set of people receives a sufficiently large enough shock to move, so a permanent utility shock raises the migration rate and population adjusts slowly. Furthermore, because the long-run elasticities are related to population shares, not migration shares, states distant from the shock adjust particularly slowly, while nearby states adjust quickly and sometimes overshoot the new long-run steady state.

Finally, the SPACE model and the moving cost model interpret the data very differently in terms of which states have become higher-utility over time. With the SPACE model, we use standard exact-hat techniques to map observed population changes onto implied utility changes across time. We can also do the same with the moving cost model. When we use U.S. population data to infer which states are gaining in terms of relative utility, we draw substantially different conclusions depending on the chosen model. This is critical if we want to estimate the role of policy or economic shocks on welfare.

We finish the paper by discussing the importance of the differences between the SPACE and moving cost model in the context of the literature. We show how the differences we highlight are central to some of the questions that are asked in the dynamic spatial literature. We discuss whether various approaches to enrich the moving cost model would deliver similar results to the SPACE model.

The specific features of the SPACE model may at first appear to be reverse engineered to match the new square root fact while lacking a basis in reality. However, we argue that the SPACE model has two very realistic features of preferences. First is that the match-specific utility for location is persistent over time. Surveys suggest that people primarily cite family and employment considerations as reasons for interstate moves (Jia, Molloy,

Smith and Wozniak, 2023).[5] People's feelings about these networks are surely correlated over time, and it is an empirical fact that each of these incentives is persistent in terms of location. Second is that match-specific utility is spatially correlated. Considering people's stated preferences, the ability to live near family is highly-correlated across space. If state i is close to family, then states near i are also close to family. Likewise, the jobs available to people in specific industries or with specific skills are geographically concentrated. Natural amenities or regional cultures—other possible sources of idiosyncratic utility—are also spatially correlated. The precise distributions we use to represent these correlations in the SPACE model are indeed convenient mathematically, but it is hard to support the argument that spatially and autocorrelated idiosyncratic utility is less realistic than the independent and identically distributed (i.i.d.) utilities of a moving cost model.

To clarify the contributions of the model, we wish to be specific regarding the difference between moving costs and persistence in match-specific utility. While mathematically straightforward to specify (as we do here), it is important to understand what each term means when mapped onto the real world. A typical moving cost model involves a one-time irreversible cost borne by people who leave one area for another. In contrast, persistent match-specific utility means that the change in utility when a person moves from one location to the other is both persistent over time and partially reversible should the person move back to the original location.[6] A moving truck and the psychological cost of throwing a goodbye party clearly are moving costs. But many factors described as "costs" in the literature are easily reversible, although the forgone benefits may decay with time. Living far from friends or a particularly amenity is a persistent and ongoing burden rather than a one-time cost. Even though such factors are often called "moving costs" in the literature, we think that terminology is used because existing models have not been able to distinguish persistence in match-specific utility from moving costs. The rest of this paper will give many reasons why this distinction is important.

[5] Koenen and Johnston (2024) documents that social networks have causal effects on migration behavior.

[6] In Appendix D, we estimate a model that includes both persistence in unobserved heterogeneity and moving costs. The model is not nearly as tractable as either model alone. The estimated model is able to match dynamics of migration even better than the SPACE model without moving costs. However, on several important dimensions, the differences between the SPACE model and the SPACE model with moving costs are quite small.

1.1 Literature

How individuals choose where to live is a classic question in the urban economics literature. Many urban models assume utility is equalized across space in the tradition of Rosen (1979) and Roback (1982). Other more quantitative models assume a discrete choice framework for locations to answer a variety of questions, such as the role of endogenous amenities on location choice (Diamond, 2016) or spatial misallocation on aggregate output (Hsieh and Moretti, 2019).

A growing share of this literature has explicitly looked at the dynamics of location choice, that is, migration. Since at least Blanchard and Katz (1992), migration has been recognized as a key feature in how regions adjust to economic shocks. In this vein, papers studying the rise or decline of regional economies place significant emphasis on migration (Caliendo, Dvorkin and Parro, 2019; Allen and Donaldson, 2020; Morris-Levenson and Prato, 2022), and especially the speed of net migration (Glaeser and Gyourko, 2005; Kleinman, Liu and Redding, 2023; Amior and Manning, 2018; Davis, Fisher and Veracierto, 2021). Similarly, when aggregating to the macroeconomic level, migration is critical to the speed of adaption to changing technologies or external shocks (Tombe and Zhu, 2019; Hao, Sun, Tombe and Zhu, 2020; Eckert and Peters, 2022; Giannone, 2017; Heise and Porzio, 2021; Bryan and Morten, 2019). A growing literature has emphasized how migration, including internal migration, plays an important role in adapting to global warming (Rudik, Lyn, Tan and Ortiz-Bobea, 2021; Cruz and Rossi-Hansberg, 2024; Oliveira and Pereda, 2020). Migration also represents an important margin when analyzing housing markets in particular (Schubert, 2021).[7] Central to many of these questions is the elasticity of local populations to different shocks over various time horizons. One of the contributions of this paper is to examine how robust those conclusions are to alternative ways of modeling migration.

Corresponding to the growth of interesting questions related to migration, there have also been advances in methods of modeling migration. Kennan and Walker (2011) formulated the canonical model of migration using the dynamic logit formulation. Kaplan and Schulhofer-Wohl (2017), Giannone et al. (2020), Porcher (2020), Mangum and Coate (2019), Zerecero (2021), and Monras (2020) have built on this formulation to incorporate additional realistic features of migration, such as richer information frictions, migrant wealth, home bias, and nested decision making.[8] Other approaches, such as Coen-Pirani

[7] Howard and Liebersohn (2021) and Howard, Liebersohn and Ozimek (2023) also study the effects of changing location choice on housing markets, but model location choice in a static discrete choice framework rather than explicitly having a notion of migration.

[8] In particular, Kaplan and Schulhofer-Wohl (2017) argues that changes in information frictions can

(2010) and Davis et al. (2021) do not use the dynamic logit framework, but have similar discrete choice models that improve the tractability in a way specific to their goals. All of these models use moving costs to explain the low rates of migration, and potentially adjust those moving costs to explain the high rates of return migration. In contrast, only one paper to our knowledge uses persistence in unobservable match-specific utility to explain low migration rates: Bayer and Juessen (2012). However, its model is too complex to tractably scale beyond two regions, limiting its use in many empirical applications.

One type of persistent match-specific utility has been modeled by Zabek (2024), Mangum and Coate (2019) and Zerecero (2021), by incorporating a preference for living in one's birthplace.[9] These models share similarities with the SPACE model, in that they also tend to feature smaller moving costs (Zerecero, 2021) and would intuitively feature more return migration than the standard moving cost model (Mangum and Coate, 2019). However, adding birthplace preferences to a moving cost model does not reproduce the square root fact and does not substantially alter many of the distinctions between the two models.

At the same time that models of internal migration have become more popular, there has also been new empirical evidence on the determinants of and constraints to migration. For example, Saks and Wozniak (2011) shows that migration is cyclical; Kleemans (2015) studies the income shocks that cause migration; Farrokhi and Jinkins (2024) examines the attachment hypothesis using a policy change among Danish refugees; Koşar, Ransom and Van der Klaauw (2021) uses a survey experiment to study how people make location decisions; and Fujiwara, Morales and Porcher (2022) proposes a methodology for uncovering information frictions in location choice. Our paper also contributes to this literature by establishing a new stylized fact that the t-year migration rate is proportional to $\sqrt{t}$.

help explain the decline in interstate migration, along with decreases in the different returns to various occupations across space. Giannone et al. (2020) builds a rich model of migration that incorporates wealth and borrowing, to analyze how credit and savings can affect if and where people choose to move. Porcher (2020) builds a tractable model of rational inattention in the dynamic migration context to argue that information frictions are one of the main reasons people do not move. Mangum and Coate (2019) models biases for birthplace and long-tenured locations, and they use their model to argue that a shift of the American population to the West and to the South is responsible for slowing labor mobility. Zerecero (2021) also examines a model that includes a preference for birthplace. Monras (2020) looks at the asymmetric response of inmigration and outmigration to local shocks, and builds a dynamic nested logit model to better capture the phenomenon.

[9]The canonical model in Kennan and Walker (2011) also includes a premium for birthplace.

2 New fact

In this section, we present the new square root fact and argue that it is a puzzle for existing models.

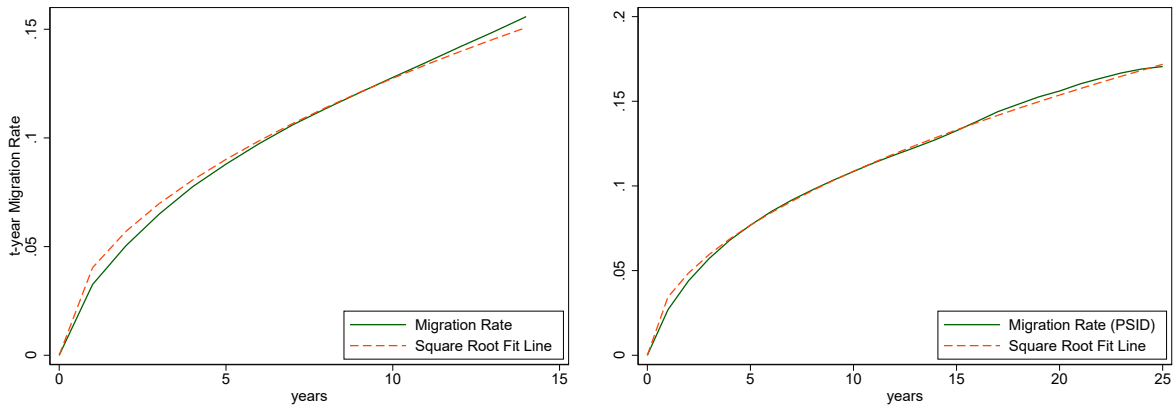
2.1 Data

Throughout the paper, we primarily measure migration using the Gies Consumer and Small Business Credit Panel (2004-2018), which is credit data from one of the leading credit report providers. It is suited to this study due to its large sample size and its panel dimension. When we can, we verify the empirical patterns using the IRS migration data, the American Community Survey, or the Panel Survey of Income Dynamics (IRS Migration Data, 2004-2018; Ruggles, Genadek, Goeken, Grover and Sobek, 2015; Panel Survey of Income Dynamics, 1969-1997).[10] The credit data is a 15-year panel of individuals making up a 1 percent random sample of the United States. It records the state of residence in each year, allowing us to calculate migration rates at longer horizons. In appendix C.1, we compare migration patterns in the GCCP and other well-known datasets, verifying the level of migration and the gravity pattern are similar. The GCCP has a one-year migration rate similar to the IRS data, but slightly higher. One contributing factor to the higher migration rate in the credit data may be that coverage of credit reports does not extend to all individuals. In particular, lower income people are less likely to have credit reports and are also less likely to move.[11]

[10]For other papers using the GCCP, see Fonseca (2022), Fonseca and Wang (2022), and Han (2023). DeWaard, Johnson and Whitaker (2019) analyzes a similar credit dataset (the Federal Reserve Bank of New York/Equifax Consumer Credit Panel) on how it can be used to study migration.

[11]While there are some well-known drawbacks to the IRS data, e.g. it is based only on tax filers, it is one of the most comprehensive administrative datasets keeping track of migration. The reasons why migration rates are particularly low in 2014 and high in 2016 are not well understood, as these anomalous values did not show up in other datasets measuring migration (see DeWaard, Hauer, Fussell, Curtis, Whitaker, McConnell, Price, Egan-Robertson, Soto and Castro (2022)).

Similarly, while credit data are not designed as a dataset to study migration, they do have location information, and the bureau gets the addresses from a person's financial accounts. The biggest concern with credit data is that moves may show up with a lag, as people do not always immediately change their addresses with their financial institutions. For our square root fact, we check the robustness to using the Panel Survey of Income Dynamics (1969-1997). The GCCP is an unbalanced panel, with yearly observations occurring in May. For matching migration patterns and rates, we focus on the 2004-2005 period, so we only observe data if they had a credit report in both of those years. For some of the dynamics, we address the unbalanced nature of the panel depending on the moments of the data that we are interested in.



(a) GCCP (b) PSID

Figure 1: Migration Rates at Different Horizons. Migration rate at year t is calculated as the percentage of people living in a different state than they did t years ago. Both datasets are unbalanced panels and use any observations in which the state of residence is recorded t years apart. Source: GCCP and PSID.

2.2 Square root fact

Define the t-year migration rate to be the share of people living in a different state than they did t years ago. The new fact is that the t-year migration rate is approximately proportional to $\sqrt{t}$. In Figure 1a, the solid line is the t-year migration rate in the GCCP. The dashed line is a constant times the square root of t, with the constant chosen to match the level of migration. As is apparent from the figure, the shape of the migration rate is very similar to the square root line.[12]

Since we cannot measure dynamic migration moments in the IRS or ACS data, one might wonder if the square root fact is driven by some sort of mismeasurement in the GCCP. In Figure 1b, we show that the square root fact is also present in data from the Panel Survey of Income Dynamics (1969-1997). In fact, when we extend the horizon to 25 years, the square root pattern holds through that longer time period as well.[13]

This fact is not the consequence of averaging across many origins and destinations or many cohorts or many years. In Appendix C.3, we show the distribution of the square root fact across pairs of origins and destinations, ages, cohorts, and starting year. For each of these figures, most of the distribution is concentrated close to the square root line.

[12]Each point is the mean of a binary variable with millions of observations, so if we tried to put standard errors on the graph, they would not be visible. Interested readers can find the standard deviation in Appendix C.2.

[13]Since mismeasurement in the GCCP may be a particular concern for young people, we show in Appendix C.4 that it holds for both people under 45 and people over 45.

We further show that there is not meaningful heterogeneity whether we look at distance between states, the time-frame of our sample, or if we focus on young or old people.

This new fact relates to the more-well-known fact that return migration is common. Many papers in the literature show a significant fraction of workers return to their previous location (e.g. Kennan and Walker, 2011; Kaplan and Schulhofer-Wohl, 2017). One consequence of this fact is that the two-year migration rate is significantly less than twice the one-year migration rate. However, we believe we are the first to document this specific relationship.

2.3 The new fact is a puzzle

The square root fact is notable not only because it is an empirical regularity in need of an explanation, but also because it is at odds with simple existing models. This section shows that the most common model of internal migration in fact leads to an approximately linear relationship between the t-year migration rate and t.

First we outline a standard moving cost migration model which we will then use to show the linear relationship. There are a continuum of individuals of mass 1, denoted by n, who can choose to live in locations denoted by i, j, or k. An agent who lived in i at $t-1$ has utility:

$$V_t(i) = \max_j \{u_{jt} - \delta_{ij} + \epsilon_{jnt} + \beta \mathbb{E} V_{t+1}(j)\} \tag{1}$$

where u_{jt} represents the common utility for everyone living in j at time t, δ_{ij} represents the bilateral moving cost between i and j, and ϵ_{jnt} represents an i.i.d. random variable with an extreme-value distribution. We assume ϵ_{jnt} has a Gumbel distribution with scale parameter 1. If we define $v_{jt} \equiv u_{jt} + \beta \mathbb{E} V_{t+1}(j)$, then the migration probability for an individual living in i is given by:

$$\frac{m_{i \to j,t}}{p_{it}} = \frac{e^{v_{jt} - \delta ij}}{\sum_k e^{v_{kt} - \delta_{ik}}} \tag{2}$$

What does this model predict for the dynamics of migration, especially for the shape of the t-year migration rate? When moving costs are high—which is required to match the low amounts of interstate migration in the data—then the following proposition shows that the t-year migration rate is approximately linear in t.

To set up the proposition, we assume that moving costs are given by $\delta_{ij} = \delta'_{ij} + \Delta$ when $i \neq j$, and $\delta_{ii} = 0$. For $i \neq j$, migration costs consist of δ'_{ij}, a pair-specific component that governs the relative amount of migration to j, and Δ, a common component which

governs the overall amount of migration in the economy. This way, when we change Δ, we are not changing the relative amount of migration from i to j versus i to k. In addition, let us define a steady-state to be when the u_{it} are constant over t, and each location's population, p_{it}, is constant over t as well.

Proposition 1. *In the steady-state of a moving cost model, as the common component of moving costs tends to infinity, the t-year migration is proportional to t.*

$$\lim_{\Delta \to \infty} \frac{m^t_{i \to j}}{m^1_{i \to j}} = t$$

where $m^t_{i \to j}$ is the t-year migration from i to j.

The proof can be found in Appendix A.1. This proposition establishes that the square root fact is not a natural consequence of our standard models. In the standard model, we infer high moving costs based on the fact that migration is low, and this proposition establishes that high moving costs imply a linear relationship between the t-year migration rate and t.

In many calibrated models, moving costs are large but less than infinity, so in Appendix C.5, we also check that the model estimated to the data would generate a relationship that looks linear. In simulations, it is also straightforward to add in complications to the model, such as including state variables for the person's location from the prior year, the person's home state, or the age of the person. In Appendix C.5, we explore whether these state variables could generate a square root relationship, but find that the relationship is still quite linear.

Of course, a flexible enough model of moving costs could match the square root pattern. For example, one could match many of the dynamic facts about migration by assuming moving costs increase with the time spent in a particular location and that moving costs are lower when returning to a past location. Similarly, if allow many different unobservable types with heterogeneity in moving costs, we could also generate the square root fact. We have four comments about this line of thought. First, these more flexible models are not the models commonly implemented in the literature. Recent papers that have estimated dynamic effects in spatial models (Caliendo et al., 2019; Kleinman et al., 2023) do not include features that would generate the square root fact. Therefore, much of what the literature knows about the macro dynamics of internal migration is built on models that do not match an important aspect of the micro dynamics.[14] Second, and related to

[14]People do write down models which have excess return migration in the first period (Kennan and

the first point, these more-flexible models lose the tractability of the standard moving cost model. If moving costs depend on a long history of locations, then calculating the macro elasticities depends on keeping track of the size of the population with each of those histories, which is computationally expensive. Third, these models can replicate any relationship between the t-year migration rate and t. To hit the square root fact, you have to parameterize these models just right to not have too much or too little concavity. The model that we propose will always approximately generate a square root, and it does so without the complexity or the flexibility of a model with lots of moving cost heterogeneity. To use the language of Fudenberg, Gao and Liang (2023), these models are not "restrictive" compared to the model presented in the next section.[15] Finally, even if the previous three points are not sufficient to convince the reader to prefer the model in this paper to a very-flexible moving cost model, remaining skeptics should still know about a new model that can fit the micro data equally well and makes different predictions for welfare and counterfactuals. This should inform us about the robustness of the conclusions drawn with the standard model.

3 A new model reconciling the square root fact

3.1 The SPACE model

This section introduces a new model of internal migration which can resolve the square root puzzle from the previous section. Rather than depend on moving costs, it assumes that the match-specific utility (the ϵ's) are spatially correlated and persistent. Because the model features Spatially and Persistently Autocorrelated Epsilons, we call it the SPACE model.

As in the moving cost model, there is a continuum of individuals with mass 1, a finite number of discrete locations, and discrete time. We keep the same notation where n denotes the individual, i the location, and t the year. Individuals choose their location to

Walker, 2011; Kaplan and Schulhofer-Wohl, 2017). But these are also not flexible enough to match the square root fact.

[15] Fudenberg et al. (2023) explains why more restrictive models are desirable compared to less restrictive: "A potential reason for this preference is that models are often meant to capture behavior in related but not identical domains. Given enough data, models that are very unrestrictive will fit any specific dataset well, but may do so by learning idiosyncratic details of those datasets that do not in fact transfer across settings. In contrast, if a highly specific and structured model happens to fit a dataset well, this may generate more confidence that the model's structure extends to other settings."

maximize utility:

$$V_t(\vec{\epsilon}_{nt}) = \max_i \{u_{it} + \epsilon_{nit}\} + \beta \mathbb{E}[V_{t+1}(\vec{\epsilon}_{nt+1})|\vec{\epsilon}_{nt}] \tag{3}$$

where u_{it} is a common flow utility for location i and ϵ_{nit} (the ith element of vector $\vec{\epsilon}_{nt}$) is a person-location-match-specific utility.[16] Note that the choice of location i does not affect the continuation value ($\mathbb{E}[V_{t+1}(\vec{\epsilon}_{t+1})|\vec{\epsilon}_t]$) because there are no moving costs, so the choice is made sequentially each period, and it has no effect on future choices. In other words, $v_{it}(\vec{\epsilon}_{nt})$, which we defined as $u_{it} + \beta \mathbb{E}[V_{t+1}(\vec{\epsilon}_{nt+1})|\vec{\epsilon}_{nt}]$, differs from u_{it} by just a constant that depends on n but not i.

To generate spatial correlation, we assume that $\epsilon_{nt} \equiv (\epsilon_{1nt}, ..., \epsilon_{Int})$ is distributed as a generalized extreme value distribution, where the marginal distribution of ϵ_{int} is a Gumbel distribution, but they are not necessarily independent of one another:[17]

$$\vec{\epsilon}_{nt} \sim F(\cdot)$$

where

$$F(\epsilon_{1nt}, ..., \epsilon_{Int}) = \exp(-G(e^{-\epsilon_{1nt}}, ..., e^{-\epsilon_{Int}})) \tag{4}$$

where G is a correlation function in the sense of McFadden (1978). To be specific, G is defined over the range of I non-negative real numbers, and it must satisfy the following properties: it is non-negative; it is homogenous of degree 1; the limit when any one of its arguments approaches infinity is infinity; and the cross-partial with respect to any k distinct arguments is nonnegative if k is odd and nonpositive if k is even.

Under these assumptions, the probability of an agent choosing location i is:

$$p_i = e^{u_i} \frac{G_i(e^{u_1}, ..., e^{u_I})}{G(e^{u_1}, ..., e^{u_I})}$$

where G_i is the partial derivative of G with respect to its ith argument. See McFadden (1978) for the derivation.

We also wish to make the ϵ correlated not just over space but also over time. To do that, we assume that the joint distribution of $\vec{\epsilon}_{nt}$ and $\vec{\epsilon}_{nt+1}$ is given by:

$$(\vec{\epsilon}_{nt}, \vec{\epsilon}_{nt+1}) \sim F_2(\cdot, \cdot)$$

[16] We do not specify the source of the u_is, so the reader can think of the SPACE model as being the migration block of a spatial model, and that the u_is would originate in the housing, production, and amenities blocks.

[17] See Lind and Ramondo (2023) as an example of a generalized extreme value distribution in trade.

where

$$F_2(\epsilon_{1nt},...,\epsilon_{Int},\epsilon_{1nt+1},...,\epsilon_{Int+1}) = \exp\big(-G\big(H(e^{-\epsilon_{1nt}},e^{-\epsilon_{1nt+1}}),...,H(e^{-\epsilon_{Int}},e^{-\epsilon_{Int+1}})\big)\big)$$

and

$$H(x_1,x_2) = \left(x_1^{\frac{1}{1-\rho}} + x_2^{\frac{1}{1-\rho}}\right)^{1-\rho}$$

where $\rho < 1$.[18] We will further assume that $G(H(\cdot,\cdot),...,H(\cdot,\cdot))$ is also a correlation function under the criteria above.[19]

Note that the cumulative distribution function of $\vec{\epsilon}_{nt}$ can be calculated by taking the limit as each element of $\vec{\epsilon}_{nt+1}$ goes to infinity. In this case, $\lim_{\epsilon_{t+1}\to\infty} H(e^{-\epsilon_t},e^{-\epsilon_{t+1}}) = e^{-\epsilon_t}$. So the marginal distribution of $\vec{\epsilon}_{nt}$ is given by F from equation (4). A similar argument applies so that the marginal distribution of $\vec{\epsilon}_{nt+1}$ is also given by F. Hence, the distribution over n of $\vec{\epsilon}_{nt}$ is time-invariant, even though an individual n's realization of $\vec{\epsilon}_{nt}$ will not be the same as $\vec{\epsilon}_{nt+1}$.

The joint distribution F_2 implies a conditional distribution $\tilde{F}(\vec{\epsilon}_{nt+1}|\vec{\epsilon}_{nt})$. This can be iterated as a Markov chain to calculate distributions of the future ϵ.

Migration occurs when the locations i that maximize $u_{it}+\epsilon_{nit}$ and $u_{i,t+1}+\epsilon_{ni,t+1}$ are different.

Proposition 2. *In the SPACE model when the u_is are fixed over time, migration from i to j is given by:*

$$m_{i\to j} = (1-\rho)e^{u_i+u_j}\left(\frac{G_i(e^{u_1},...,e^{u_I})G_j(e^{u_1},...,e^{u_I})}{G(e^{u_1},...,e^{u_I})^2} - \frac{G_{ij}(e^{u_1},...,e^{u_I})}{G(e^{u_1},...,e^{u_I})}\right) \tag{5}$$

where G_{ij} is the second derivative of G with respect to its ith and jth elements.

[18] ρ is not the correlation between ϵ_{int} and ϵ_{int+1}. The correlation is $1-(1-\rho)^2$ (Cascetta, 2001).

[19] Suppose G has a cross-nested structure as follows:

$$G(x_1,...,x_I) = \sum_k \left(\sum_{i\in I_k} \alpha_{ik} x_i^{\frac{1}{1-\gamma_k}}\right)^{1-\gamma_k}$$

for nests k and weights $\alpha_{ik} > 0$. Each I_k is a subset of the set of locations $\{1,...,I\}$. Then, as long as $\rho > \gamma_k$ for all k, G will be a correlation function. This can be shown because G is the sum of several nested logit correlation functions, and the sum of correlation functions is a correlation function. This claim is important because then the function F_2 is indeed a proper cumulative distribution function. Lind and Ramondo (2023) prove that cross-nested formulations of G can approximate any correlation function, so combined with this proposition, then any correlation function can be approximated by a correlation function that permits some ρ that induces the persistence we desire to add to the model.

Alternatively,

$$m_{i\to j} = (1-\rho)p_i p_j \left(1 + \tau_{ij}(e^{u_1}, ..., e^{u_I})\right) \tag{6}$$

where $\tau_{ij} = -\frac{G_{ij}(e^{u_1},...,e^{u_I})G(e^{u_1},...,e^{u_I})}{G_i(e^{u_1},...,e^{u_I})G_j(e^{u_1},...,e^{u_I})}$.

Please refer to Appendix A.2 for the proof. Equation (6) resembles a gravity equation. τ_{ij}, which is non-negative, is related to the correlation between two alternatives. More correlated locations—in the sense that they have higher τ_{ij}—will have more migration between them. Conversely, if migration is more correlated over time (higher ρ), there will be less migration.

Corollary 1. *The cross-elasticity of population to utility is given by the migration rate times a constant.*

$$\frac{\partial p_i}{\partial u_j} = -\frac{1}{1-\rho} m_{i\to j}$$

when $i \neq j$.

See Appendix A.3 for the proof. This corollary is important because in many applications, we are interested in the population elasticity to local shocks. This corollary tells us that migration is a sufficient statistic to know these elasticities up to a constant.

The migration in Corollary 1 is a steady-state migration between i and j, which in the steady-state of the model is equal to the migration from j to i. In the data, these two migration rates are highly related, even conditional on origin, destination, and distance (see Appendix C.6). However, migration is rarely exactly balanced in the data. In subsequent parts of the paper, we will adjust the data to be balanced by using the logarithmic average i.e. $(m_{i\to j,t} - m_{j\to i,t})/\log(m_{i\to j,t}/m_{j\to i,t})$, because in a calibrated parametric version of the model that we introduce in the next section, this is a precise approximation of the steady-state migration we need for Corollary 1 (see Appendix B.4 for details).

Having set up the SPACE model, we now present a proposition analogous to Proposition 1 from the moving cost model, to see whether the SPACE model can match the square root fact. In this case, the limit we consider is for the persistence parameter ρ to approach 1.

When ρ is close to 1, ϵ_{int} will resemble a random-walk with logistic innovations. It is well known that the standard deviation of a random walk grows with the square root of time. As the standard deviation grows, the odds of crossing a threshold—which corresponds to moving in this model—grow roughly proportionally to the standard deviation, hence generating the square root fact. We formalize this intuition in the following lemma:

Lemma 1. *Define Λ^x to be the sum of x i.i.d. mean-zero logistic random variables. Then in steady-state of the SPACE model,*

$$\lim_{\rho \to 1} \frac{m^t_{i \to j}}{m^1_{i \to j}} = \frac{\mathbb{E}[|\Lambda^{2t}|]}{\mathbb{E}[|\Lambda^2|]}$$

where $\mathbb{E}[|\cdot|]$ denotes the mean absolute deviation (MAD).

Please refer to Appendix A.4 for the proof. This proposition establishes that for ρ close to 1, relative migration over t years will be proportional to the mean absolute deviation of the sum of $2t$ independent logistic random variables, following the intuition from above. Importantly, we can calculate bounds on this ratio:

Proposition 3. *In the steady-state of the SPACE model, as $\rho \to 1$, the ratio of t-year migration to 1-year migration for any state-pair is bounded below by $\sqrt{t}$ and bounded above by $\sqrt{\pi/3}\sqrt{t}$:*

$$\sqrt{t} \leq \lim_{\rho \to 1} \frac{m^t_{i \to j}}{m^1_{i \to j}} \leq \sqrt{\frac{\pi}{3}}\sqrt{t}$$

Please refer to Appendix A.5 for the proof. This proposition shows the SPACE model is able to naturally match the square root fact—albeit approximately. There are several senses in which the fit is only approximate. First, the proposition only establishes bounds. Fortunately, $\sqrt{\pi/3}$ is approximately 1.023, so the bound is tight.[20] Second, the proposition relies on $\rho \to 1$, while in practice we will calibrate a $\rho < 1$. Our calibration in the next section calibrates a high persistence, and we check the square root fact quantitatively under that calibration. Finally, the proposition assumes the economy is in steady-state, whereas in the data, the utility of each state (u_i) is changing. We check that the changes are not large enough to change the square root fact in simulations in Appendix B.4.

The SPACE model has a closed-form solution for aggregate welfare. Aggregating over the ϵ's, welfare is given by:

$$W_t = \mathbb{E}_{\vec{\epsilon}_{nt}}[V_t(\vec{\epsilon}_{nt})] = \sum_{s=0}^{\infty} \beta^s \left(\bar{\gamma} + \log G(\exp(u_{1t+s}), ..., \exp(u_{It+s}))\right)$$

[20] A tight band in Proposition 3 is intuitive because, if the innovations were normally distributed and Lemma 1 referenced normal distributions instead of logistic ones, the ratio would be exactly the square root, without the 2.3 percent deviation. Similarly, if the lemma were based on the standard deviation rather than the mean absolute deviation, the ratio would also be precisely the square root. Although this exact relationship does not hold for the MAD of logistic distributions, logistics are approximately normal, and the MAD is approximately proportional to the standard deviation for many distribution families. Consequently, it is not surprising that the relationship holds approximately.

where $\bar{\gamma}$ is the Euler-Mascheroni constant. This follows directly from McFadden (1978). It follows that

$$\frac{\partial W_t}{\partial u_{it+s}} = \beta^s p_{is} \qquad \text{and} \qquad \frac{\partial^2 W_t}{\partial u_{it+s} \partial u_{jt+s}} = -\beta^s \frac{1}{1-\rho} m_{i \to j,s}$$

where $m_{i \to j,s}$ represents the amount of steady-state migration from i to j that would occur when the location utilities are given by u_{is}.[21] So to second-order, the welfare effects of changes in local utility are given by population and migration, adjusted for parameters β and ρ.

3.2 Calibration

Many functions G will exactly match the populations and migration in the data. And in many settings where the outcome of interest is the change in populations, the choice of a specific G makes no first-order difference, due to the previous result that $\frac{\partial p_i}{\partial u_j} = \frac{1}{1-\rho} m_{i \to j}$.

Nonetheless, in some applications, it will be helpful to take a stand on a specific functional form for G. It is sometimes useful to have a functional form that allows for analytical formulas for population changes and the ability to simulate an individual's migration path.[22] We will also be interested in finding a G that is consistent with a value of ρ near 1, because that is the parameterization of ρ that leads to the square root fact.

Consider the following G functional form, which generates a cross-nested logit with nests that can contain any number of locations:

$$G(x_1, ..., x_I) = \sum_{q=1}^{Q} \left(\sum_{i \in \mathcal{N}_q} (w_{iq} x_i)^{\frac{1}{1-\gamma}} \right)^{1-\gamma}, \tag{7}$$

where nests are indexed by $q = 1, \ldots, Q$, each nest q corresponds to a subset $\mathcal{N}_q \subseteq \{1, \ldots, I\}$, and w_{iq} are nonnegative weights that we will calibrate to match populations and migration (with $w_{iq} = 0$ if $i \notin \mathcal{N}_q$). This choice of functional form leads to a cross-nested logit, in which nests can contain an arbitrary number of locations. The within-nest elasticity of substitution is $\frac{1}{1-\gamma}$ and the across-nest elasticity of substitution is 1.

In order to do individual-level simulations, it will be helpful to parameterize γ and the nest weights w_{iq} as a function of ρ, and consider the limit as $\rho \to 1$. In particular, we

[21] A good approximation of the this migration would be the logarithmic mean of $m_{i \to j}$ and $m_{j \to i}$ in the data (see Appendix B.4).

[22] Another minor benefit to a calibration is a closed-form solution for migration outside of steady-state (see Appendix B.4).

will parameterize

$$1-\gamma = \frac{1-\rho}{1-\tilde{\rho}}$$

where $\tilde{\rho}$ is fixed, so as $\rho \to 1$, $\gamma \to 1$ as well. This will allow us to consider the nest for each person as fixed over time, but allow for agents to move within nests. In order to approximately match the square root fact, we will aim to choose a $\tilde{\rho}$ as close to 1 as possible while also matching migration rates in the data.

Rather than calibrating the w_{iq}, it is more intuitive to calibrate

$$\tilde{w}_{iq} \equiv (w_{iq}e^{u_i})^{\frac{1}{1-\gamma}} \left(\sum_{k \in \mathcal{N}_q} (w_{kq}e^{u_k})^{\frac{1}{1-\gamma}} \right)^{-\gamma}$$

Note that these $\tilde{w}$ are also a function of ρ since they depend on γ. From there, the w's can be backed out as $w_{iq} = e^{-u_i}\, \tilde{w}_{iq}^{1-\gamma} \left(\sum_{k \in \mathcal{N}_q} \tilde{w}_{kq} \right)^{\gamma}$. One intuition is that these adjusted weights represent the mass of people that choose a location–nest pair: $\tilde{w}_{iq}$ choose i via nest q. Under the above definitions,

$$p_i = \sum_{q:\, i \in \mathcal{N}_q} \tilde{w}_{iq} \tag{8}$$

and

$$m_{i \to j} = (1-\tilde{\rho}) \sum_{q:\, \{i,j\} \subseteq \mathcal{N}_q} \frac{\tilde{w}_{iq}\tilde{w}_{jq}}{\sum_{k \in \mathcal{N}_q} \tilde{w}_{kq}}. \tag{9}$$

See Appendix B for the derivation.

This calibration is useful for two reasons. First, it leads to a simple formula for populations and straightforward exact-hat algebra. Define $\hat{p}_i = p_i'/p_i$, the growth in population. Similarly, define $\hat{u}_i = \exp(\frac{u_i' - u_i}{1-\rho})$, the change in local utility, normalized such that the elasticity of location choice with respect to utility does not diverge as $\rho \to 1$. Then, the change in populations as a function of the change in utility can be expressed as

$$\hat{p}_i = \hat{u}_i^{1-\tilde{\rho}} \sum_{q=1}^{Q} \frac{\tilde{w}_{iq}}{p_i} \frac{\sum_{k \in \mathcal{N}_q} \tilde{w}_{kq}}{\sum_{k \in \mathcal{N}_q} \tilde{w}_{kq} \hat{u}_k^{1-\tilde{\rho}}} \tag{10}$$

See Appendix B for the derivation.

Second, this calibration is useful because each person is drawn into choosing between $|\mathcal{N}_q|$ locations and never considers anywhere else. Furthermore, within a nest, one can simulate the $|\mathcal{N}_q|$ locations as independent of each other, since the nesting structure

accounts for all of the correlation. This is helpful because drawing from high-dimensional generalized extreme value distributions is computationally-intensive. By only having to draw extreme value distributions that are correlated from one period to the next, the computational burden is greatly alleviated.

Of course, there are still many ways to choose the $\tilde{w}_{iq}$ and $\tilde{\rho}$. A natural way might be to maximize $\tilde{\rho}$ as to match the square root fact, while treating equations (8) and (9) as constraints. Unfortunately, this is not a problem on which we could make significant progress. However, if we restrict $\mathcal{N}_q$ to be sets of locations with two elements, the problem becomes tractable. This restriction has the added advantage of only having to simulate a correlated GEV for two locations at a time, making the computational burden of doing simulations lighter. In particular, the solution to this constrained optimization problem is that the $1-\tilde{\rho}$ is the largest eigenvalue of an $I \times I$ matrix M, defined by

$$M_{ij} = \begin{cases} \frac{m_{i\to j}}{p_i} & \text{if } i \neq j \\ \frac{m_i}{p_i} & \text{if } i = j \end{cases}$$

where $m_i = \sum_{k\neq i} m_{i\to k}$.[23] The calibrated $\tilde{w}_{ij} = \frac{1}{1-\tilde{\rho}} m_{i\to j}\,(1 + \ell_j/\ell_i)$, where ℓ_i is the ith element of the corresponding eigenvector, where we shorten the notation so that $\tilde{w}_{ij}$ is the share of people that choose i from the nest that includes i and j, i.e. $\tilde{w}_{ij} \equiv \tilde{w}_{iq}$ where $q = \{i, j\}$. The calibrated value of $\tilde{\rho}$ is 0.892, which, as we will show in the next section, is close enough to 1 to generate the square root fact in simulations.[24] Details on this calibration can be found in Appendix B.[25]

3.3 Dynamics of migration in the calibrated SPACE model

Under the calibration from the previous section, each agents' actions are as if they were drawing two Gumbel-distributed ϵ's that are independent between the two locations, but correlated over time with correlation parameter $\tilde{\rho}$. Compared to a generalized extreme value distribution with many dimensions, it is computationally efficient to simulate se-

[23] For the calibration, we use the migration matrix summing all one-year migration from 2004-2018 in the GCCP, and using the logarithmic average to symmetrize it. (See Appendix B.4 for the rationale of using the logarithmic average.)

[24] The autocorrelation implied by this parameter is $1 - (1-\tilde{\rho})^2 = 0.988$.

[25] The downside of assuming two elements per nest is that it means that simulations will never have anyone that chooses to live in more than two locations over time, whereas in the data people move to their third or fourth state somewhat commonly. In Appendix B.2, we propose a calibration that has a reasonably high $\tilde{\rho}$ and which would generate people moving to many potential states.

quences of conditional Gumbels where ϵ_{it} and ϵ_{it+1} have joint distribution

$$\exp(-(e^{-\epsilon_{it}/(1-\tilde{\rho})} + e^{-\epsilon_{it+1}/(1-\tilde{\rho})})^{1-\tilde{\rho}})$$

because there is a closed-form formulation of the CDF of ϵ_{it+1} given ϵ_{it}. We can easily draw a sequence for location i and another sequence for location j, and then that person will choose i when $\epsilon_{it} > \epsilon_{jt} + \log \frac{\tilde{w}_{ji}}{\tilde{w}_{ij}}$.

In this section, we simulate 100,000 agents using the above algorithm for each state-pair, and record their location choice for 15 years.[26] Each observation is weighted by $\tilde{w}_{ij} + \tilde{w}_{ji}$ to make the simulation representative, and we can use the simulated panel to measure dynamic moments of migration.

In Figure 2a, we show that the t-year migration rate does follow a square root pattern in both the data and in a simulation of the SPACE model. The model does not match the data perfectly, with the lines diverging over time. In part, this is because the model is calibrated to match the one-year migration rate, and the 14-year migration rate, which in the simulation is about $\sqrt{14}$ times the one-year migration rate, is sensitive to that choice. This figure also presents the same exercise for a simulation of the moving cost model, which is much more linear and diverges much more from the data.[27,28]

Of course, the t-year migration rate is not the typical way the dynamic moments of migration are presented in the data, so it is worth examining whether the model captures more commonly examined moments. A natural moment is the conditional probability of moving given a previous move, i.e. the hazard rate of migration.

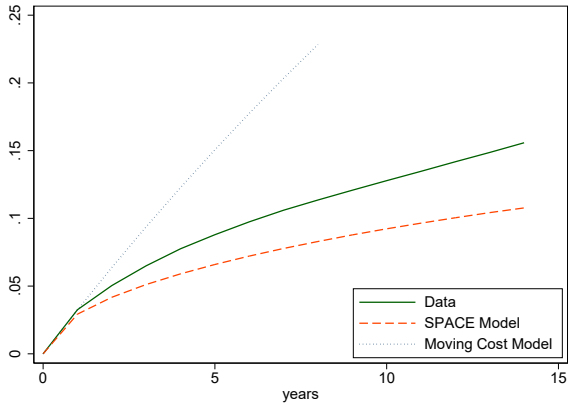
Figure 2b shows the probability of another migration at different time horizons after an interstate move.[29] Since we do not target these statistics in the calibration, the simulated statistics do not match the data perfectly. But the general pattern is similar, especially its decay as the person has lived in the state for longer.

[26]The ϵ_{it} draws are reused for each state-pair, but the threshold for which state to live in changes by state-pair.

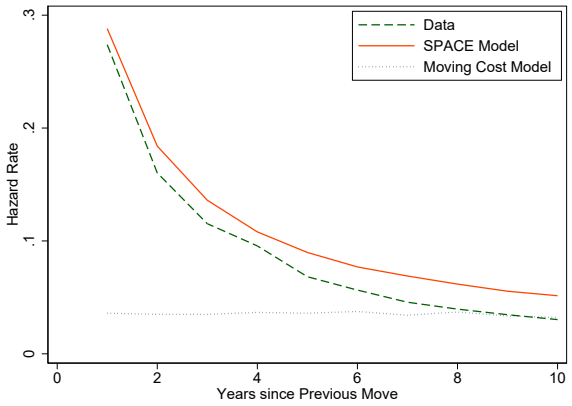
[27]For the moving cost model simulation, we use the observed 1-year migration probabilities for the entire GCCP panel, 2004-2018, and simulate 15 years of location choices for 300 million people.

[28]A model that features both persistent unobserved heterogeneity and moving costs can improve upon the fit of all three dynamics moments mentioned in this section. Intuitively, because moving costs and persistence generate different patterns for the t-year migration rate, they can be separately identified off of short-run and long-run migration rates (e.g. 1-year and 10-year migration). However, such a model is much less tractable than either the SPACE model or the moving cost model. In practice, a model that combines both features calibrates significant persistence and small moving costs, and leaving the moving costs out of the model completely makes little quantitative difference. See Appendix D for details.

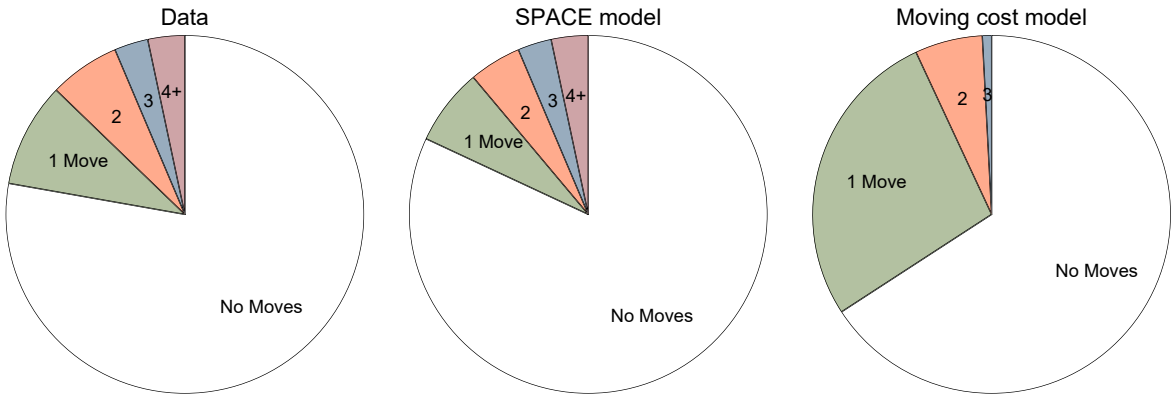
[29]To be included in this analysis, a person must show up for the number of years that would be necessary to calculate the statistic, but we do not use a balanced panel.



(a) Migration rate in the data and models



(b) Hazard rate of migration



(c) Number of moves in 14 years

Figure 2: Dynamic moments. In panel (a), the t-year migration rate is calculated as the percent of people living in a different state than they were t years ago. Data are from an unbalanced panel, and included any observations from 2004-2018 for which the state of residence is observed t years apart. In panel (b), the conditional probability of migration is plotted. For the value at x years, the probability of migration is conditional on the person having migrated x years previously and remained in the same state ever since. In panel (c), the number of moves in 14 years is calculated for people whose state is observed in every year from 2004-2018.

The intuition for the decreasing hazard rate over time is simple. Conditional on having moved recently, the agents are likely relatively indifferent between the two regions and are likely to move back. The longer they have stayed in one region, the more likely that their accumulated utility shocks have drawn them further away from being indifferent, so the probability of migration decreases over time. The literature has typically focused on the concept of "attachment" to explain this phenomenon (Mangum and Coate, 2019; Farrokhi and Jinkins, 2024). In the SPACE model, people who have lived in a location for longer are more attached, but it is because their repeated decision not to move has revealed that they like the location, not due to an economic force that increases their utility by staying there longer.

In addition to hazard rates, another readily observable statistic is the distribution of the number of interstate moves over time. Figure 2c looks at how many moves are made over a 14 year period. In the data, a large majority of people make zero moves, but some people make many moves. Here, we include in this chart only people for whom we have data in all 15 years (for up to 14 possible moves). The model is able to capture the large fraction of people that never move. It also comes close to the data on the number of people that move once or twice. Importantly, it captures the fact that a few percent of people move four or more times over the 14 years. The figure also includes similar statistics for a moving cost model, which fits the data substantially less well.

Together, these moments reinforce that the SPACE model captures essential features of migration dynamics in the data, especially compared to the standard moving cost model.

4 Does the new model matter?

The previous section introduced a new model and showed that it did a better job at matching dynamic moments in the micro data. In this section, we explore the implications of the new model by comparing it to the workhorse model. For some questions, we find the differences are minimal, while for others we find substantial differences.

4.1 Micro forecasting accuracy

An important distinction between the two models is their implications for forecasting individuals' locations. Suppose we observe the agent's location in 2004 and wish to forecast where they will live in every subsequent year until 2018. We use the calibrated

versions of each model to perform the forecasting. We judge the performance of the models using the mean Kullback-Leibler divergence. Specifically, we use the same simulations as before, and then for each initial location, we calculate the simulated probability that a person who was in state i in 2004 ends up in state j in year t.[30] Then using people's true locations in the data, we calculate the Kullback-Liebler divergence:

$$KL_t = \frac{1}{N} \sum_n \log \left(\frac{\mathbb{P}_{\text{data}}(\text{lives in } j \text{ in } t | \text{lived in } i \text{ in 2004})}{\mathbb{P}_{\text{model}}(\text{lives in } j \text{ in } t | \text{lived in } i \text{ in 2004})} \right)$$

for each year. We plot KL_t in Figure 3.

Both models initially have a low Kullback-Leibler divergence, as they are about equally good at predicting locations in 2005 since they were both parameterized to match the migration data in that year. But over time, the moving cost model's Kullback-Leibler divergence grows sharply, suggesting that the log likelihood is on average about 0.12 log-points worse per observation than the maximum possible performance of any model by 2018. In contrast, the SPACE model has a Kullback-Leibler divergence of 0.014 log-points in 2018, suggesting limited room for further improvement.

The SPACE model provides more accurate forecasts over longer horizons because it matches the dynamic moments. In particular, many more people end up moving away from their initial state in the moving cost model because moving probabilities are independent over time, whereas the SPACE model is better able to match the total number of people who leave.

4.2 Moving costs need not be large

Kennan and Walker (2011) estimate an average moving cost of $312,146 (in 2010 dollars).[31] This estimate exceeds six times the median household income in that year, which was $49,445 (Census, 2011).[32] Such a large cost is one of the main reasons that most people do not move, in their model. Economists can argue about whether that number is reasonable, and even within Kennan and Walker (2011), there is substantial heterogeneity in moving costs. In contrast, the SPACE model can match the main facts about internal migration without any moving costs. In other words, the fact that most people do not move is not sufficient evidence to conclude that moving costs are large.

[30] This exercise is why we chose to simulate hundreds of millions of agents, as we had to ensure that every origin-destination-timeframe that shows up in the data is also in the simulations.

[31] They also include an analysis of moving costs conditional on moving, but they include the payoff shocks in the moving costs, and consequently find that the average moving cost is actually very negative.

[32] Many papers estimate moving costs of a similar magnitude. See Table 2 of Howard (2026).

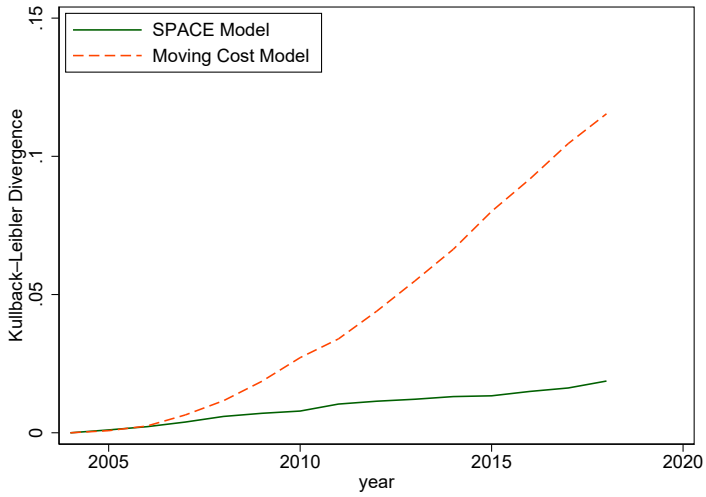


Figure 3: Kullback-Leibler divergence of the SPACE and Moving Cost models. For the SPACE model, 100,000 agents per state-pair are simulated over 15 years, and KL_t is calculated based on the simulated data using the formula in Section 4.1, weighted to be nationally representative. For the moving cost model, 300,000,000 observations are simulated over 15 years, and KL_t is calculated based on the simulation. The data is from the GCCP.

In Appendix D, we calibrate an augmented version of the SPACE model that also features moving costs to match the one-year and 10-year rates of migration. The moving costs are about two orders of magnitude smaller than had we calibrated a model that had only moving costs.

A common counterfactual in the literature is to consider changes in moving costs (Kennan and Walker, 2011; Schubert, 2021; Zerecero, 2021), which is also an actual policy used by some localities.[33] For example, Kennan and Walker (2011) finds that a moving subsidy could substantially increase the gross migration rate. In a moving cost model, a temporary incentive to move to location i has a very persistent impact on the population of i. In contrast, in our model, a moving subsidy would encourage people to relocate, but only for as long as the subsidy lasts. After the subsidy expires, they are no more likely to remain in the place they moved to than they would be to live there had the subsidy never occurred.[34]

One particular way of lowering "moving costs" may be improving infrastructure such as

[33]A handful of cities around the United States offer monetary incentives to relocate (Cornerstone Home Lending, 2021).

[34]One could imagine other economic reasons that populations remain higher after a population expansion, such as the accumulation of housing capital or agglomeration in that place.

roads, which increases migration (Morten and Oliveira, 2024). The SPACE model would interpret the increased migration as an increase in the correlation of idiosyncratic utility across the locations that the infrastructure connected.[35] In the moving cost model, the increased migration must reflect lower moving costs, and so welfare would have increased, net the cost of the roads. In the modified SPACE model, though, the welfare effects are much more muted.[36]

4.3 Macro population elasticities

Another common use for migration models is to calculate population elasticities to changes in a location's utility, v_i, both in the short run and the long run. In the standard moving cost model, the elasticity of population with respect to v_j is given by:

$$\frac{\partial \log p_i}{\partial v_j} = -\sum_k \frac{m_{k\to i}}{p_i}\frac{m_{k\to j}}{p_k}$$

This is approximately proportional to the migration rate between i and j, since the two terms where $k = i$ or $k = j$ are much larger than the remaining terms.

The elasticities in the two models are similar in that they are approximately proportional to the migration rate when migration rates are low. Given that the scale of v_j is not specified in either model, the constant terms are ignorable without loss of generality. However, in the long-run, the similarities of population elasticities between the SPACE model and the moving cost model break down. Consider a one-time permanent change in v_{it} for the SPACE model or the moving cost model. In the SPACE model, the population elasticity is still exactly the same, since it is given by corollary 1. This is not the case with the moving cost model. In fact, when migration is symmetric and moving costs are high, we show that the long-run elasticities converge to a static logit.

Proposition 4. *In the long-run steady state of a moving cost model, if migration is symmetric, in the sense that* $\lim_{\Delta\to\infty} m_{i\to j}/m_{j\to i} = 1$*, then*

$$\lim_{\Delta\to\infty} \frac{\partial \log p_i}{\partial v_j} = \begin{cases} -2p_j & \text{if } i \neq j \\ 2 - 2p_j & \text{if } i = j \end{cases} \tag{11}$$

[35] This increased correlation is a natural consequence of infrastructure that reduces travel time, as it makes social networks and amenities in the other city more accessible, thereby increasing the correlation of the idiosyncratic utility experienced in the newly connected places. Hence, new infrastructure can increase migration even if the infrastructure itself is not the primary means of migration.

[36] Of course, other welfare benefits, such as those that come through increased trade, are not contingent on the migration model.

See Appendix A.6 for the proof. Recall that the total population is mass 1, so p_i is both the population of i and its population share. These steady-state elasticities are the same as a static logit, and a key difference from the SPACE model is that they have no relationship to migration data. In Appendix C.8, we show equation (11) is a good numerical approximation to the long-run of a calibrated moving cost model, when moving costs lead to realistic migration rates, rather than being infinite.

In the long-run, the moving cost model has no notion that closer states are better substitutes or that states with higher migration are likely to be more impacted by a change in the other state. The moving cost model does not imply that a state with a high migration rate will exhibit a greater long-run elasticity than a state with a low migration rate. Rather, the only pertinent factor is the population share of the state receiving the shock to calculate all the relevant elasticities (approximately).[37]

Figure 4 summarizes the conclusions of this section. In the short-run (Panel a), the SPACE model and the moving cost model predict practically the same cross-elasticities of population. But the in the long-run (Panel b), there is almost no relationship between the two.

4.4 Macro adjustment dynamics

Given the differences in long-run population elasticities, it follows that the intermediate dynamics must also be different across the two models. We illustrate by considering a one-time permanent shock to Louisiana utility, $v_{\text{Louisiana}}$, to see how populations respond in each model. In Figure 5 panels (a) and (b), we show that the population dynamics after a one-time permanent shock are starkly different.[38]

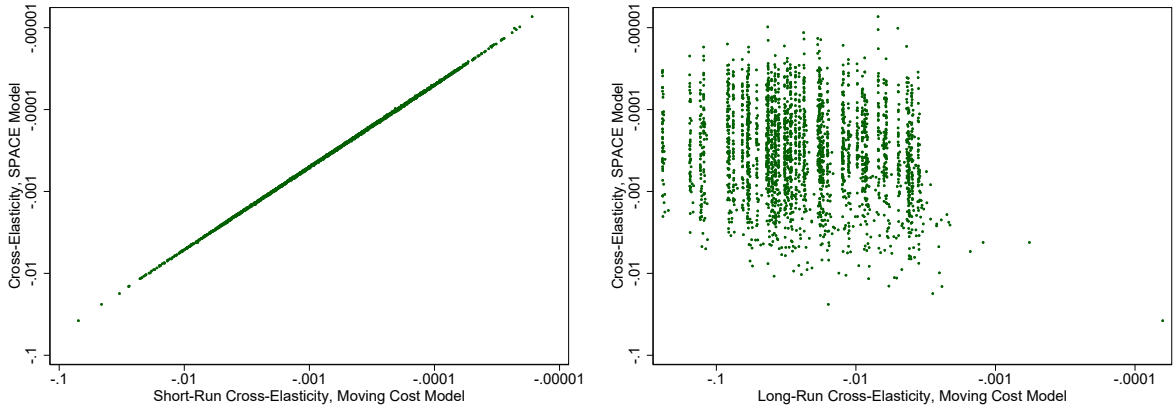
In the SPACE model (Panel a), the population adjustment in Louisiana is immediate, and the population stops adjusting after the first period. In contrast, in the moving cost model (Panel b), the population adjustment takes many years, with the model finally

[37] In the moving cost model, the migration rates govern the speed of adjustment (Kleinman et al., 2023), but not the long-run effects.

Including birthplace as a state variable in the moving cost model would mean that adjustments would depend on the population shares of individuals by birthplace. To this extent, migration between states would be correlated to the population cross-elasticities (Zabek, 2024).

Other features of the model can also change the population elasticities. For example, Monte, Redding and Rossi-Hansberg (2018) adds commuting to a static model of location choice to generate variation in population elasticities.

[38] We chose Louisiana because in Appendix C.9, we do further analysis on the Hurricane Katrina episode. The size of the shocks in each model is normalized to have a long-term effect of about 5 percent of the population for Louisiana. However, the scale is irrelevant to the focus of this exercise, which is the dynamics.



(a) Short-run (b) Long-run

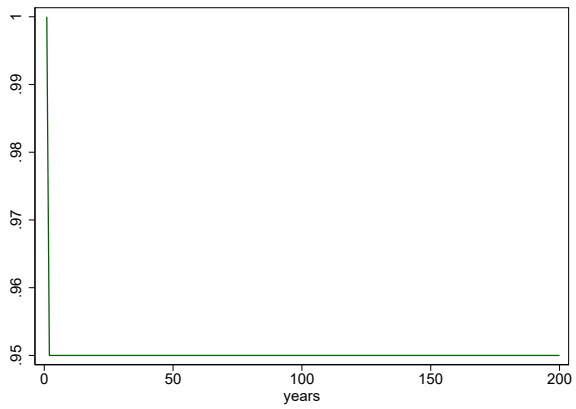
Figure 4: A comparison of the population cross-elasticities between the SPACE and moving cost models. For both figures, each dot represents a pair of states. The point is located at the population cross-elasticity between the two states in each of the two models. The constant multiplicative terms are ignored, since each model is subject to a normalization of utility. Data for migration and population shares comes from the GCCP. Both axes in both graphs have log scales.

approaching a steady-state after almost 200 years.[39]

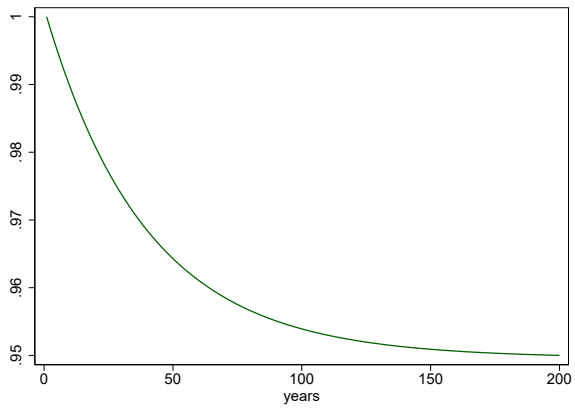
In Panels (c) and (d) we illustrate the dynamics for other states in response to the same shock to Louisiana's utility. In the SPACE model (Panel c), there is a bigger population effect on Mississippi than there is on New York, as one would expect due to the geography. Again, the adjustment occurs immediately. But in Panel (d), the dynamics follow interesting and perhaps counterintuitive patterns. In New York, the population adjustment is particularly slow because of low migration between Louisiana and New York. In contrast, for Mississippi, the population dramatically overshoots its long-run steady-state because there is so much migration between Louisiana and Mississippi.

This exercise is intended to highlight difference between the models, not argue in favor of one over the other. While the time series of Louisiana's population share after Hurricane Katrina does resemble that of the SPACE model (see Figure A10a), we do not wish to judge whether Katrina represented a one-time permanent disutility shock. Nonetheless, the differences in the transition dynamics imply that the models will interpret empirical facts differently. We will elaborate on this point in Section 5.1.

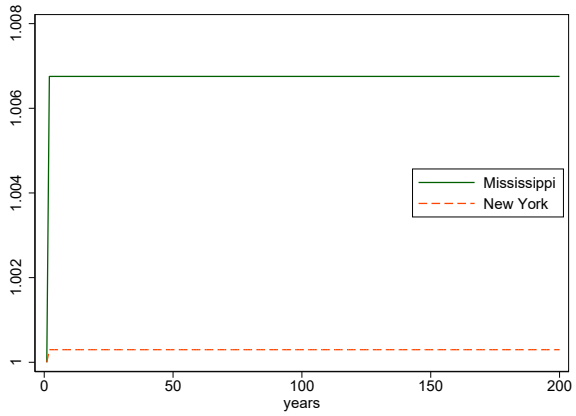
[39]In an estimated model with both persistent unobserved heterogeneity and moving costs, we still see a very fast population adjustment (see Appendix D). It is much closer to the macro dynamics of the SPACE model than the moving cost model.



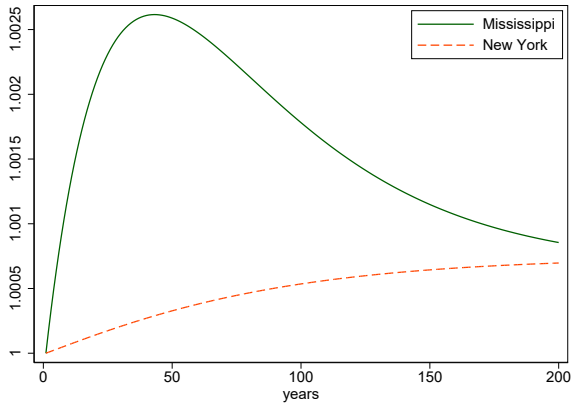
(a) Louisiana, SPACE model



(b) Louisiana, moving cost model



(c) Mississippi and New York, SPACE model



(d) Mississippi and New York, moving cost model

Figure 5: Population Dynamics after a one-time permanent change in $v_{\text{Louisiana}}$, in the SPACE model and the moving cost model, for Louisiana, Mississippi, and New York. The shock size is chosen to cause a long-run five percent decline in Louisiana's population. The y-axis in each figure measures the population in each state, relative to the population in that state at time $t = 0$, before the shock. Mississippi and New York were chosen to represent two states for which there is high gross migration with Louisiana, and low gross migration with Louisiana, respectively.

4.5 Implied utility changes

Given the differences in population elasticities, it must be the case that the models will imply different things about changes in utility over time. This is important for papers that wish to estimate the welfare effect of some policy or event that varies across space. The population changes in the SPACE model can be expressed with exact hat algebra. Under our previous calibration, the population changes are given by:

$$\hat{p}_i = \hat{v}_i^{1-\tilde{\rho}} \sum_{j \neq i} \frac{\tilde{w}_{ij}}{p_i} \frac{\tilde{w}_{ij} + \tilde{w}_{ji}}{\tilde{w}_{ij}\hat{v}_i^{1-\tilde{\rho}} + \tilde{w}_{ji}\hat{v}_j^{1-\tilde{\rho}}}$$

Recall that $\hat{v}_i = \hat{u}_i$ due the continuation value differing from the current value by only a constant. This equation can be numerically inverted to calculate the $\hat{v}_i$ based on the $\hat{p}_i$. Note that while this exact formula depends on our calibration, it will give the same answer to first-order of any choice of G, since the population elasticities to v are pinned down by migration rates.

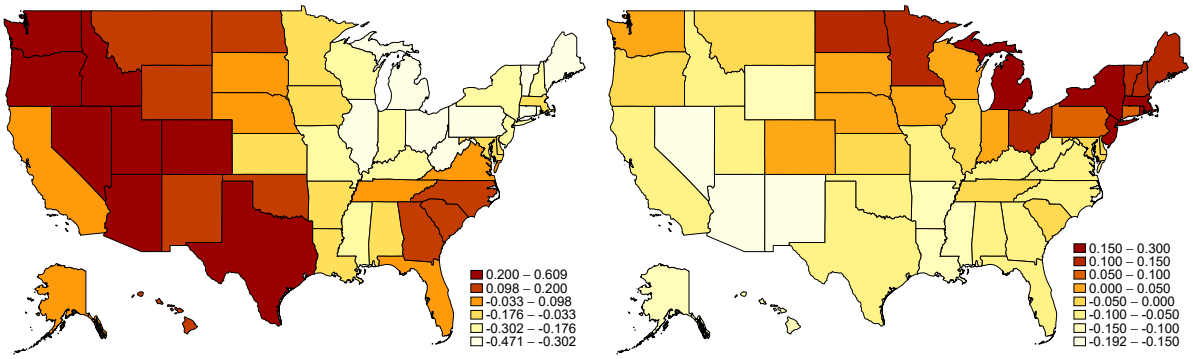
For the moving cost model, the exact hat algebra is given by:

$$\hat{p}_{it} = \hat{v}_i \sum_j \frac{\frac{m_{j \to i}}{p_i} \hat{p}_{jt-1}}{\sum_k \frac{m_{j \to k}}{p_j} \hat{v}_k}$$

This formula is similar to the dynamic hat algebra derived by Caliendo et al. (2019) (see Appendix B.3 for the derivation). Note that it depends on $\hat{p}_{it-1}$ as well as $\hat{p}_{it}$. For this exact hat algebra, we need to know about the change in how populations are changing, rather than simply the change in population. Again, this is straightforward to invert numerically, given data on $\hat{p}_{it}$ and $\hat{p}_{it-1}$.

To illustrate this, we consider the utility changes implied by the SPACE model and moving cost model from 2005-2018, the span of our data. We show the results in Figure 6. In the SPACE model, the places with the largest increases in relative utility are in the South and West, places that have seen large growth in population. New England and the Rust Belt have some of the largest relative decreases. In the moving cost model, the utility changes are almost the opposite. New York and New England have increased in relative utility, while the South and the West have mostly had relative declines.[40] Overall, the correlation between the log of utility changes implied by the two models is -0.497. This has important implications for estimating spatial models. For example, if one wanted

[40]The utility changes in the moving costs model are driven by changes in the net migration rate. While the South and West have had higher net migration rates over much of this period, the rate of that growth has slowed, leading us to infer declining utility.



(a) SPACE model (b) Moving cost model

Figure 6: Change in utilities v_j, 2005-2018, implied by the SPACE model and the moving cost model. Implied utility changes are calculated by inverting the exact hat algebra in Section 4.5. Population changes are from the Bureau of Economic Analysis (2004-2018).

to estimate the effects of a wage, rent, or amenity change on utility, you would get very different answers using the implied utilities from the SPACE model versus a moving cost model.

5 Discussion

5.1 Relation to literature

Given the major differences between the two models on several of these questions, it is worth emphasizing why these differences matter. For the micro forecasting and for the interpretation, the reasons to care about differences are readily apparent, but for the population elasticities and the dynamics, it is important to consider the context of the literature.

One big question in the literature is to what extent does population adjust to shocks? For example, if one particular location has a shock that permanently increases the utility of living there, how will that affect the distribution of the population around the country? This is a question that is asked by Caliendo et al. (2019) with respect to the China shock of Autor, Dorn and Hanson (2013), by Giannone (2017) with respect to skill-biased technological change, and by Cruz and Rossi-Hansberg (2024), Oliveira and Pereda (2020), and Rudik et al. (2021) with respect to climate change.[41] In the short run, both

[41] A reader may wonder why we do not replicate one of these papers to highlight the differences. However, doing such a replication would erroneously indicate that the SPACE model is less good at hitting the medium-run dynamics of population adjustment. The reason for this is that even if these

models agree that places with high gross migration, such as D.C., have highly elastic population responses to local shocks, compared to places with little migration. However, in the long-run, the SPACE model continues to make this prediction, while the moving cost model predicts similar elasticities for all locations. Similarly, in the short-run, both models agree that the population effects are felt in states that have lots of migration between them and the state with the shock. A shock to D.C. will affect Maryland and Virginia more in the short run than it will affect Arizona. This is consistent with the "donut" phenomenon during the recent COVID-19 crisis, as areas around major cities have experienced population and house price growth in recent years (Ramani and Bloom, 2021). Again, this holds in the long-run too for the SPACE model, but migration would not generate a long-run donut phenomenon in the moving cost model.

Another key question in this literature is how quickly the migration adjustment takes place (Kleinman et al., 2023; Amior and Manning, 2018; Caliendo et al., 2019). In the data, migration is sometimes quite persistent. For example, the Rust Belt has had low inmigration for decades, and the Sun Belt has had high inmigration for decades. In the moving cost model, much of this persistence is due to the fact that migration is inherently persistent (Kleinman et al., 2023). For example, the moving cost model might infer that the Rust Belt had a large negative utility shock a long time ago and the process of moving out has been very slow.

The SPACE model interprets this fact as being about the utility of a location adjusting slowly. It could be that the underlying shocks to utility are slow. Another possibility is that there was a large initial shock, but some equilibrium mechanism causes utility to decline slowly. For example, housing is durable, and so housing becomes cheap as people move out, keeping utility from falling too quickly (Glaeser and Gyourko, 2005). In Appendix E, we show the effects of a productivity shock to Louisiana in a model that has SPACE migration and durable housing. Indeed, the population adjustments are slow once housing is introduced. Similarly, there may be similar mechanisms through the labor market that make structural transformation slow.[42] The SPACE model would emphasize

models are not explicitly targeting the medium-run dynamics, they do get to choose what features of the world to add and can choose to include or not include features that will get the dynamics right. For example, Glaeser and Gyourko (2005) argues that the reason declining cities decline slowly is because the housing stock is slow to depreciate. Many of the quantitative papers do not have this feature. Therefore, replacing the migration block from those models would lead to unrealistic dynamics, if we did not also add in a feature like the one in Glaeser and Gyourko (2005). A reader may think of the moving costs as representing these other features, such as housing depreciation or labor market frictions. In that case, it would be preferable to explicitly model those features. For example, we add a model of housing depreciation in Appendix E.

[42]Kleinman et al. (2023) discuss how having location specific durable capital can keep wages high after

these forces as explanations for persistence in migration, whereas the moving cost model would attribute the persistence to an inherent property of migration itself, and find less explanatory power for these forces.

Conversely, some adjustments in the data are quite fast, such as Hurricane Katrina. The SPACE model can easily accommodate these fast adjustments by assuming a fast adjustment in utility. Even in a model with housing, utility can adjust quickly because a hurricane destroys much of the housing stock, forcing people out immediately. In contrast, the moving cost model requires implausible assumptions regarding utility to generate the fast decline in population followed by a small rebound the next year. The moving cost model implies that utility in Louisiana two years after Hurricane Katrina is higher than it was before the hurricane (see Appendix C.9).

Other papers are concerned with the effects of location-specific shocks on aggregate outcomes such as welfare or output (Tombe and Zhu, 2019; Eckert and Peters, 2022; Hsieh and Moretti, 2019). The degree to which people are able to move is an important factor for these outcomes. Because the second order effects are determined by these population elasticities, it shows that a shock that affects a higher-gross-migration place will have larger total effects on welfare if the shock is positive, and smaller effects if the shock is negative. Again, this holds in the short-run for both models, but in the long-run only for the SPACE model.

Finally, the spatial correlation of a shock is an important determinant of its welfare consequences. If a negative shock is extremely localized, it may be easy to move away from it, and there will be lots of insurance. If shocks are correlated across space, then the welfare effects may be much less insurable. Of course, in the long-run of a moving cost model, this effect will no longer hold.

5.2 Richer moving cost models

In Section 4, we compared the SPACE model to a simple version of a moving cost model. Yet as mentioned in the literature review, there is significant research that enriches the moving cost model to match a variety of facts (Kaplan and Schulhofer-Wohl, 2017; Giannone et al., 2020; Porcher, 2020; Mangum and Coate, 2019; Zerecero, 2021; Monras, 2020). Kennan and Walker (2011) includes features to increase home bias and return migration. As we demonstrated in Section 2.3, neither return migration, home bias, nor age is a sufficient feature to match the square root fact. So the puzzle that motivated

a negative productivity shock.

the model is not dependent on us having considered a simple version of the moving cost model. Nonetheless, in this section we ask whether the differences we highlighted between the moving cost model and the SPACE model are robust to considering richer versions of the moving cost model. In the rest of this section, we consider each of the differences we highlighted before.

For the prediction of individuals' locations, the reason that the SPACE model outperformed the moving cost model was because it could hit the square root fact. So if extensions of the moving cost model still do not hit the square root fact, they might be an improvement at predicting locations, but are not going to make the same predictions as the SPACE model.

For the interpretation of why people rarely move, some of the additional features imply lower estimated moving costs (Zerecero, 2021; Giannone et al., 2020), but never by orders of magnitude. So the difference between the SPACE model—which has no moving costs—and any moving cost model will remain substantial.

For population elasticities, more complex moving cost models feature long-run elasticities which may not necessarily be the same as a static logit model. For example, models with home bias have more similar elasticities to the SPACE model than the baseline moving cost model does.[43] However, except by coincidence, none of the additional features would generate the property that the long-run and short-run population elasticities are the same. So the marked difference between the SPACE model and the moving cost model will remain, both for long-run elasticities and for adjustment speeds.

For the implied utility changes, the exact implied utilities will certainly change with a richer model. However, it does not change the fundamental fact that the implied utility changes are primarily correlated to changes in migration in the moving cost model, and primarily correlated to changes in populations in the SPACE model.

6 Conclusion

This paper documents a new empirical regularity in internal migration: the t-year migration rate scales approximately with the square root of t. This pattern poses a challenge to the standard moving cost framework, which implies a linear relationship under com-

[43] With home bias, the elasticities are given by $\lim_{\Delta \to \infty} \partial \log p_j / \partial v_k = -\sum_i w_{ij} p_{ik}$, where $w_{ij} = \frac{p_{ij}}{p_j}$ is the share of people in j who are from i, and p_{ik} is the share of people from i living in k. These population shares are likely correlated to the amount of bilateral migration. However, it is a different formula, and the population share levels are likely different than the migration rates, so the dynamics in the two models will still be different.

monly used assumptions. To account for this fact, we propose a new model of location choice, the SPACE model, that introduces persistent and spatially correlated unobserved heterogeneity. The model is analytically tractable, matches key features of the data including the square root relationship and return migration rates, and provides a better out-of-sample forecast of location dynamics than the canonical moving cost alternative.

The SPACE model makes substantially different predictions from existing models. The SPACE model generates population elasticities that are proportional to bilateral migration flows, leading to markedly different long-run responses to local shocks than a standard moving cost model, as well as different dynamics. It also changes conclusions about the size of moving costs and which parts of the country have become better places to live.

The tractability of the SPACE model, especially its closed-form expressions for population and compatibility with exact hat algebra, makes it a practical foundation for embedding into broader spatial equilibrium frameworks. It is our hope that that future researchers will incorporate the SPACE framework into more complex models, much as researchers today use the moving cost framework in models that also include trade frictions and capital accumulation. We also hope that the SPACE model will be used in static models of location choice to generate realistic cross-elasticities of populations that are disciplined by migration data.

References

Allen, Treb and Dave Donaldson, "Persistence and path dependence in the spatial economy," 2020. National Bureau of Economic Research Working Paper.

Amior, Michael and Alan Manning, "The persistence of local joblessness," *American Economic Review*, 2018, *108* (7), 1942–70.

Autor, David, David Dorn, and Gordon H Hanson, "The China syndrome: Local labor market effects of import competition in the United States," *American Economic Review*, 2013, *103* (6), 2121–68.

Bair, Jacques, Piotr Błaszczyk, Robert Ely, Valérie Henry, Vladimir Kanovei, Karin U. Katz, Mikhail G. Katz, Semen S. Kutateladze, Thomas McGaffey, David M. Schaps, David Sherry, and Steven Shnider, "Is Mathematical History Written by the Victors?," *Notices of the American Mathematical Society*, 2013, *60* (7), 886–904. Published July/August 2013.

Baum-Snow, Nathaniel and Lu Han, "The microgeography of housing supply," *Journal of Political Economy*, 2024, *132* (6), 1897–1946.

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.

Bilal, Adrien and Esteban Rossi-Hansberg, "Anticipating climate change across the united states," Technical Report, National Bureau of Economic Research 2023.

Blanchard, Olivier Jean and Lawrence F Katz, "Regional evolutions," *Brookings Papers on Economic Activity*, 1992, *1992* (1), 1–75.

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

Bureau of Economic Analysis, "Regional Economic Accounts, Annual Personal Income by State," 2004-2018. https://apps.bea.gov/regional/downloadzip.htm.

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.

Cascetta, Ennio, *Transportation Systems Engineering: Theory and Methods*, Vol. 49 of *Applied Optimization*, New York: Springer Science & Business Media, 2001.

Census, U.S., "Income, Poverty and Health Insurance Coverage in the United States: 2010," 2011.

Coen-Pirani, Daniele, "Understanding gross worker flows across US states," *Journal of Monetary Economics*, 2010, *57* (7), 769–784.

Correia, Sergio, Paulo Guimaraes, and Thomas Zylkin, "PPMLHDFE: Stata module for Poisson pseudo-likelihood regression with multiple levels of fixed effects," 2019.

Cruz, José-Luis and Esteban Rossi-Hansberg, "The economic geography of global warming," *Review of Economic Studies*, 2024, *91* (2), 899–939.

Davis, Morris A, Jonas DM Fisher, and Marcelo Veracierto, "Migration and urban economic dynamics," *Journal of Economic Dynamics and Control*, 2021, *133*, 104234.

Deryugina, Tatyana, Laura Kawano, and Steven D. Levitt, "The Economic Impact of Hurricane Katrina on Its Victims," *American Economic Journal: Applied Economics*, 2018, *10* (2), 202–233.

DeWaard, Jack, Janna Johnson, and Stephan Whitaker, "Internal migration in the United States: A comprehensive comparative assessment of the Consumer Credit Panel," *Demographic research*, 2019, *41*, 953.

_ , Mathew Hauer, Elizabeth Fussell, Katherine J Curtis, Stephan D Whitaker, Kathryn McConnell, Kobie Price, David Egan-Robertson, Michael Soto, and Catalina Anampa Castro, "User beware: Concerning findings from the post 2011–2012 US internal revenue service migration data," 2022.

Diamond, Rebecca, "The determinants and welfare implications of US workers' diverging location choices by skill: 1980-2000," *American Economic Review*, 2016, *106* (3), 479–524.

Eckert, Fabian and Michael Peters, "Spatial Structural Change," 2022. NBER Working Paper.

Farrokhi, Farid and David Jinkins, "Root growing and path dependence in location choice: Evidence from Danish refugee placement," *Regional Science and Urban Economics*, 2024, *105*, 103975.

Fonseca, Julia, "Less Mainstream Credit, More Payday Borrowing? Evidence from Debt Collection Restrictions," *Journal of Finance*, 2022.

_ and Jialan Wang, "How Much do Small Businesses Rely on Personal Credit?," 2022.

Frobenius, Ferdinand Georg, "Über Matrizen aus nicht negativen Elementen," *Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften*, 1912, pp. 456–477.

Fudenberg, Drew, Wayne Gao, and Annie Liang, "How flexible is that functional form? Quantifying the restrictiveness of theories," 2023.

Fujiwara, Thomas, Eduardo Morales, and Charly Porcher, "A Revealed Preference Approach to Measuring Information Frictions in Migration Decisions," 2022.

Giannone, Elisa, "Skill-Biased Technical Change and Regional Convergence," 2017.

_ , Qi Li, Nuno Paixao, and Xinle Pang, "Unpacking Moving," 2020.

Gies Consumer and Small Business Credit Panel, 2004-2018. Accessed July 20, 2021.

Glaeser, Edward L and Joseph Gyourko, "Urban decline and durable housing," *Journal of Political Economy*, 2005, *113* (2), 345–375.

Groen, Jeffrey A., Mark J. Kutzbach, and Anne E. Polivka, "Storms and Jobs: The Effect of Hurricanes on Individuals' Employment and Earnings over the Long Term," *Journal of Labor Economics*, 2020, *38* (3), 653–685.

Han, Peter, "The Portfolio Lending Premium in the Mortgage Market," 2023.

Hao, Tongtong, Ruiqi Sun, Trevor Tombe, and Xiaodong Zhu, "The effect of migration policy on growth, structural change, and regional inequality in China," *Journal of Monetary Economics*, 2020, *113*, 112–134.

Heise, Sebastian and Tommaso Porzio, "The aggregate and distributional effects of spatial frictions," 2021. National Bureau of Economic Research Working Paper.

Howard, Greg, "JUE Insight: Moving Cost Magnitudes in Moving Cost Models," *Journal of Urban Economics*, 2026.

_ and Jack Liebersohn, "Why is the rent so darn high? The role of growing demand to live in housing-supply-inelastic cities," *Journal of Urban Economics*, 2021, *124*, 103369.

_ , _ , and Adam Ozimek, "The Short-and Long-Run Effects of Remote Work on US Housing Markets," *Journal of Financial Economics*, 2023.

Hsieh, Chang-Tai and Enrico Moretti, "Housing constraints and spatial misallocation," *American Economic Journal: Macroeconomics*, 2019, *11* (2), 1–39.

IRS Migration Data, "Internal Revenue Service Statistics of Income Tax Statistics - Migration Data," 2004-2018. https://www.irs.gov/statistics/soi-tax-stats-migration-data.

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.

Kaplan, Greg and Sam Schulhofer-Wohl, "Understanding the long-run decline in interstate migration," *International Economic Review*, 2017, *58* (1), 57–94.

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

Kleemans, Marieke, "Migration choice under risk and liquidity constraints," 2015.

Kleinman, Benny, Ernest Liu, and Stephen Redding, "Dynamic spatial general equilibrium," *Econometrica*, 2023.

Koenen, Martin and Drew Johnston, "Social ties and residential choice: Micro evidence and equilibrium implications," Technical Report, Technical report, Harvard University 2024.

Koşar, Gizem, Tyler Ransom, and Wilbert Van der Klaauw, "Understanding migration aversion using elicited counterfactual choice probabilities," *Journal of Econometrics*, 2021.

Lending, Inc. Cornerstone Home, "These 11 U.S. Cities Will Pay You to Live There," 2021. Accessed August 9, 2022. https://www.houseloanblog.net/get-paid-to-relocate/.

Lind, Nelson and Natalia Ramondo, "Trade with Correlation," *American Economic Review*, February 2023, *113* (2), 317–53.

Mangum, Kyle and Patrick Coate, "Fast locations and slowing labor mobility," 2019.

McFadden, Daniel, "Modelling the choice of residential location," 1978.

Missouri Data Census Center, "Geocorr 2000: Geographic Correspondence Engine," 2000. `https://mcdc.missouri.edu/applications/geocorr2000.html`.

Molloy, Raven, Christopher L Smith, and Abigail Wozniak, "Internal migration in the United States," *Journal of Economic Perspectives*, 2011, *25* (3), 173–96.

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

Monte, Ferdinando, Stephen J Redding, and Esteban Rossi-Hansberg, "Commuting, migration, and local employment elasticities," *American Economic Review*, 2018, *108* (12), 3855–90.

Moretti, Enrico, "Real wage inequality," *American Economic Journal: Applied Economics*, 2013, *5* (1), 65–103.

Morris-Levenson, Joshua and Marta Prato, "The Origins of Regional Specialization." PhD dissertation, The University of Chicago 2022.

Morten, Melanie and Jaqueline Oliveira, "The effects of roads on trade and migration: Evidence from a planned capital city," *American Economic Journal: Applied Economics*, 2024, *16* (2), 389–421.

Oliveira, Jaqueline and Paula Pereda, "The impact of climate change on internal migration in Brazil," *Journal of Environmental Economics and Management*, 2020, *103*, 102340.

Panel Survey of Income Dynamics, 1969-1997. https://psidonline.isr.umich.edu/.

Perron, Oskar, "Zur Theorie der Matrizen," *Mathematische Annalen*, 1907, *64* (2), 248–263.

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

Ramani, Arjun and Nicholas Bloom, "The Donut effect of COVID-19 on cities," 2021. National Bureau of Economic Research Working Paper.

Roback, Jennifer, "Wages, rents, and the quality of life," *Journal of Political Economy*, 1982, *90* (6), 1257–1278.

Rosen, Sherwin, "Wage-based indexes of urban quality of life," *Current Issues in Urban Economics*, 1979, pp. 74–104.

Rudik, Ivan, Gary Lyn, Weiliang Tan, and Ariel Ortiz-Bobea, "The Economic Effects of Climate Change in Dynamic Spatial Equilibrium," 2021.

Ruggles, Steven, Katie Genadek, Ronald Goeken, Josiah Grover, and Matthew Sobek, "Integrated Public Use Microdata Series: Version 6.0 [dataset]," `http://doi.org/10.18128/D010.V6.0` 2015. Minneapolis: University of Minnesota.

Saks, Raven E and Abigail Wozniak, "Labor reallocation over the business cycle: New evidence from internal migration," *Journal of Labor Economics*, 2011, *29* (4), 697–739.

Schubert, Gregor, "House Price Contagion and US City Migration Networks," 2021.

Silva, JMC Santos and Silvana Tenreyro, "The log of gravity," *The Review of Economics and Statistics*, 2006, *88* (4), 641–658.

Tombe, Trevor and Xiaodong Zhu, "Trade, migration, and productivity: A quantitative analysis of china," *American Economic Review*, 2019, *109* (5), 1843–72.

Tonelli, Leonida, "Sull'integrazione per parti," *Annali di Matematica Pura ed Applicata*, 1909, *17*, 97–180.

Tuggle, Laura, "Affordable Housing in the Wake of Katrina." PhD dissertation, Tulane University 2010. Ph.D. Dissertation.

Vincenty, Thaddeus, "Direct and inverse solutions of geodesics on the ellipsoid with application of nested equations," *Survey review*, 1975, *23* (176), 88–93.

Zabek, Michael A, "Local ties in spatial equilibrium," *American Economic Journal: Macroeconomics*, 2024.

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

Online Appendix

A Proofs

A.1 Proof of Proposition 1

We aim to prove:

> In the steady-state of a moving cost model, as the common component of moving costs tends to infinity, the t-year migration is proportional to t.
>
> $$\lim_{\Delta \to \infty} \frac{m^t_{i \to j}}{m^1_{i \to j}} = t$$
>
> where $m^t_{i \to j}$ is the t-year migration from i to j.

Proof: In the steady-state of the moving cost model, consider the probability of someone living in i living in j after t years. This probability is given by:

$$\frac{m^t_{i \to j}}{p_i} = \sum_{k_1, k_2, k_3, ..., k_{t-1} \in I} \frac{m_{i \to k_1}}{p_i} \left(\prod_{s=1}^{t-2} \frac{m_{k_s \to k_{s+1}}}{p_{k_s}} \right) \frac{m_{k_{t-1} \to j}}{p_{k_{t-1}}}$$

which is the sum of probabilities over all possible paths between i and j over t years.

Consider

$$\lim_{\Delta \to \infty} \frac{m^t_{i \to j}}{m_{i \to j}} = \lim_{\Delta \to \infty} \sum_{k_1, k_2, k_3, ..., k_{t-1} \in I} \frac{\frac{m_{i \to k_1}}{p_i} \left(\prod_{s=1}^{t-2} \frac{m_{k_s \to k_{s+1}}}{p_{k_s}} \right) \frac{m_{k_{t-1} \to j}}{p_{k_{t-1}}}}{\frac{m_{i \to j}}{p_i}}$$

We can calculate each term in the summation. First, consider all the summations such that there exists a cutoff $T \in \{1, 2, 3, ..., t\}$ such that for all $s < T$, $k_s = i$ and for all $s \geq T$, $k_s = j$. This corresponds to all the paths from i to j that involve only one move. There are t such combinations, corresponding to each possible period the person makes a move. For each combination, the product in the numerator is of migration probabilities from i to i (non-migration), one migration probability from i to j, and migration probabilities from j to j:

$$\frac{\left(\prod_{s=1}^{T-1} \frac{m_{i \to i}}{p_i} \right) \frac{m_{i \to j}}{p_i} \left(\prod_{s=T+1}^{t} \frac{m_{j \to j}}{p_j} \right)}{\frac{m_{i \to j}}{p_i}}$$

For the non-migration probabilities, the limit as $\Delta \to \infty$ of $\frac{m_{i \to i}}{p_i}$ or $\frac{m_{j \to j}}{p_j}$ is 1. The

remaining term, $\frac{m_{i\to j}}{p_i}$ cancels with the denominator, so the limit is 1.

Next consider all other terms, which will involve more than one move. For each, there is one year in which the person moves away from i and another year in which they move from their next location to somewhere else. Put these two terms first in the product:

$$\frac{m_{i\to \ell}/p_i}{m_{i\to j}/p_i} \cdot \frac{m_{\ell\to k_s}}{p_\ell} \cdot ...$$

The first fraction is $\frac{\exp(v_{\ell t}-\delta_{i\ell})}{\exp(v_{jt}-\delta_{ij})}$ regardless of Δ, which is a constant. The second term converges to zero when $\Delta \to \infty$. All the following terms are bounded by 0 and 1, so the whole product converges to 0.

Therefore, the sum is over t 1's and many zeros. Hence,

$$\lim_{\Delta\to\infty} \frac{m_{i\to j,t}}{m_{i\to j}} = t$$

□

A.2 Proof of Proposition 2

We aim to prove:

> In the SPACE model when the u_is are fixed over time, migration from i to j is given by:
>
> $$m_{i\to j} = (1-\rho)e^{u_i+u_j}\left(\frac{G_i(e^{u_1},...,e^{u_I})G_j(e^{u_1},...,e^{u_I})}{G(e^{u_1},...,e^{u_I})^2} - \frac{G_{ij}(e^{u_1},...,e^{u_I})}{G(e^{u_1},...,e^{u_I})}\right)$$
>
> where G_{ij} is the second derivative of G with respect to its ith and jth elements.
>
> Alternatively,
>
> $$m_{i\to j} = (1-\rho)p_i p_j\left(1+\tau_{ij}(e^{u_1},...,e^{u_I})\right)$$
>
> where $\tau_{ij} = -\frac{G_{ij}(e^{u_1},...,e^{u_I})G(e^{u_1},...,e^{u_I})}{G_i(e^{u_1},...,e^{u_I})G_j(e^{u_1},...,e^{u_I})}$.

Proof: The definition of migration is

$$P(u_i+\epsilon_{it} \geq \max_k u_k+\epsilon_{kt}, u_j+\epsilon_{jt+1} \geq \max_k u_k+\epsilon_{kt+1})$$

when $(\epsilon, \epsilon_{t+1}) \sim F_2$. We can express this as an integral:

$$\int_{\substack{u_i+\epsilon_{it}\geq \max_k u_k+\epsilon_{kt} \\ u_j+\epsilon_{jt+1}\geq \max_k u_k+\epsilon_{kt+1}}} dF_2(\vec{\epsilon}_t, \vec{\epsilon}_{t+1})$$

where dF_2 is equal to the $2I$th cross-partial derivative of F_2 with respect to each of its arguments.

We first integrate over all the ϵ's except ϵ_{it} and ϵ_{jt+1}. For all the ϵ_{kt} at time t, the limits of integration are $(-\infty, u_i + \epsilon_{it} - u_k)$, and at time $t+1$, the limits are $(-\infty, u_j + \epsilon_{jt+1} - u_k)$. The integration leaves us with the cross-partial derivative of F_2 with respect to the remaining two ϵ_{it} and ϵ_{jt+1}, evaluated at $u_i + \epsilon_{it} - u_k$ for all the ϵ_{kt} or $u_j + \epsilon_{jt+1} - u_k$ for all the ϵ_{kt+1}.

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{\partial^2 F_2}{\partial \epsilon_{it}\partial \epsilon_{jt+1}}(u_i - u_1 + \epsilon_{it}, ..., u_i - u_I + \epsilon_{it}, u_j - u_1 + \epsilon_{jt+1}, ..., u_j - u_I + \epsilon_{jt+1}) d\epsilon_{it} d\epsilon_{jt+1}$$

Recall the definition of F_2:

$$F_2 = \exp(-G((e^{-\epsilon_{1t}/(1-\rho)} + e^{-\epsilon_{1t+1}/(1-\rho)})^{1-\rho}, ..., (e^{-\epsilon_{It}/(1-\rho)} + e^{-\epsilon_{It+1}/(1-\rho)})^{1-\rho})))$$

The derivative of F_2 with respect to ϵ_{it} is:

$$\frac{\partial F_2}{\partial \epsilon_{it}} = \exp(-G) G_i (e^{-\epsilon_{it}/(1-\rho)} + e^{-\epsilon_{it+1}/(1-\rho)})^{-\rho} e^{-\epsilon_{it}/(1-\rho)}$$

And the cross-partial derivative with respect to ϵ_{it} and ϵ_{jt+1} is:

$$\begin{aligned}\frac{\partial^2 F_2}{\partial \epsilon_{it}\partial \epsilon_{jt+1}} =& \exp(-G) G_i G_j \cdot (e^{-\epsilon_{it}/(1-\rho)} + e^{-\epsilon_{it+1}/(1-\rho)})^{-\rho} e^{-\epsilon_{it}/(1-\rho)} (e^{-\epsilon_{jt}/(1-\rho)} + e^{-\epsilon_{jt+1}/(1-\rho)})^{-\rho} e^{-\epsilon_{jt+1}/(1-\rho)} \\ &- \exp(-G) G_{ij} \cdot (e^{-\epsilon_{it}/(1-\rho)} + e^{-\epsilon_{it+1}/(1-\rho)})^{-\rho} e^{-\epsilon_{it}/(1-\rho)} (e^{-\epsilon_{jt}/(1-\rho)} + e^{-\epsilon_{jt+1}/(1-\rho)})^{-\rho} e^{-\epsilon_{jt+1}/(1-\rho)}\end{aligned}$$

G is homogeneous of degree 1, so G_i is homogeneous of degree 0, and G_{ij} is homogeneous of degree -1. Therefore,

$$\begin{aligned}G((e^{-(u_i - u_1 + \epsilon_{it})/(1-\rho)} + e^{-(u_j - u_1 + \epsilon_{jt+1})/(1-\rho)})^{1-\rho}, ..., (e^{-(u_i - u_I + \epsilon_{it})/(1-\rho)} + e^{-(u_j - u_I + \epsilon_{jt+1})/(1-\rho)})^{1-\rho}) \\ = G(e^{u_1}, ..., e^{u_I})(e^{-(u_i + \epsilon_{it})/(1-\rho)} + e^{-(u_j + \epsilon_{jt+1})/(1-\rho)})^{1-\rho}\end{aligned}$$

and

$$\begin{aligned}G_i((e^{-(u_i - u_1 + \epsilon_{it})/(1-\rho)} + e^{-(u_j - u_1 + \epsilon_{jt+1})/(1-\rho)})^{1-\rho}, ..., (e^{-(u_i - u_I + \epsilon_{it})/(1-\rho)} + e^{-(u_j - u_I + \epsilon_{jt+1})/(1-\rho)})^{1-\rho}) \\ = G_i(e^{u_1}, ..., e^{u_I})\end{aligned}$$

and

$$G_{ij}((e^{-(u_i-u_1+\epsilon_{it})/(1-\rho)}+e^{-(u_j-u_1+\epsilon_{jt+1})/(1-\rho)})^{1-\rho},...,(e^{-(u_i-u_I+\epsilon_{it})/(1-\rho)}+e^{-(u_j-u_I+\epsilon_{jt+1})/(1-\rho)})^{1-\rho})$$
$$= \frac{G_{ij}(e^{u_1},...,e^{u_I})}{(e^{-(u_i+\epsilon_{it})/(1-\rho)}+e^{-(u_j+\epsilon_{jt+1})/(1-\rho)})^{1-\rho}}$$

$G(e^{u_1},...,e^{u_I})$ is a constant; let us denote it by $\bar{G}$. Similarly, denote $\bar{G}_i \equiv G_i(e^{u_1},...,e^{u_I})$ and $\bar{G}_{ij} \equiv G_{ij}(e^{u_1},...,e^{u_I})$.

The integral becomes

$$\begin{aligned}
&\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp(-(e^{-(u_i+\epsilon_{it})/(1-\rho)}+e^{-(u_j+\epsilon_{jt+1})/(1-\rho)})^{1-\rho}\bar{G})\bar{G}_i\bar{G}_j \\
&\quad\times (e^{-\epsilon_{it}/(1-\rho)}+e^{-(\epsilon_{jt+1}+u_j-u_i)/(1-\rho)})^{-\rho}e^{-\epsilon_{it}/(1-\rho)}(e^{-(\epsilon_{it}+u_i-u_j)/(1-\rho)}+e^{-\epsilon_{jt+1}/(1-\rho)})^{-\rho}e^{-\epsilon_{jt+1}/(1-\rho)} \\
&\quad- \exp(-(e^{-(u_i+\epsilon_{it})/(1-\rho)}+e^{-(u_j+\epsilon_{jt+1})/(1-\rho)})^{1-\rho}\bar{G})\bar{G}_{ij} \\
&\quad\times \frac{1}{(e^{-(u_i+\epsilon_{it})/(1-\rho)}+e^{-(u_j+\epsilon_{jt+1})/(1-\rho)})^{1-\rho}} \\
&\quad\times (e^{-\epsilon_{it}/(1-\rho)}+e^{-\epsilon_{it+1}/(1-\rho)})^{-\rho}e^{-\epsilon_{it}/(1-\rho)}(e^{-\epsilon_{jt}/(1-\rho)}+e^{-\epsilon_{jt+1}/(1-\rho)})^{-\rho}e^{-\epsilon_{jt+1}/(1-\rho)}d\epsilon_{it}d\epsilon_{jt+1}
\end{aligned}$$

This can be simplified:

$$\begin{aligned}
&\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp(-(e^{-(u_i+\epsilon_{it})/(1-\rho)}+e^{-(u_j+\epsilon_{jt+1})/(1-\rho)})^{1-\rho}\bar{G}) \\
&\quad\times \left((\bar{G}_i\bar{G}_j - \bar{G}_{ij}\frac{1}{(e^{-(u_i+\epsilon_{it})/(1-\rho)}+e^{-(u_j+\epsilon_{jt+1})/(1-\rho)})^{1-\rho}}\right) \\
&\quad\times e^{u_i+u_j}(e^{-(\epsilon_{it}+u_i)/(1-\rho)}+e^{-(\epsilon_{jt+1}+u_j)/(1-\rho)})^{-2\rho}e^{-(\epsilon_{it}+u_i)/(1-\rho)}e^{-(\epsilon_{jt+1}+u_j)(1-\rho)}d\epsilon_{it}d\epsilon_{jt+1}
\end{aligned}$$

Consider the following u-substitution:

$$\begin{aligned}
v &= (e^{-(u_i+\epsilon_{it})/(1-\rho)}+e^{-(u_j+\epsilon_{jt+1})/(1-\rho)})^{1-\rho} \\
dv &= -(e^{-(u_i+\epsilon_{it})/(1-\rho)}+e^{-(u_j+\epsilon_{jt+1})/(1-\rho)})^{-\rho}e^{-(u_j+\epsilon_{jt+1})/(1-\rho)}d\epsilon_{jt+1}
\end{aligned}$$

With this u-substitution, the integral simplifies considerably:

$$\int_{-\infty}^{\infty}\int_{e^{-(u_i+\epsilon_{it})}}^{\infty} \exp(-\bar{G}v)\left((\bar{G}_i\bar{G}_j - \bar{G}_{ij}\frac{1}{v}\right)e^{u_i+u_j}v^{-\rho/(1-\rho)}e^{-(\epsilon_{it}+u_i)/(1-\rho)}dvd\epsilon_{it}$$

The integrand is non-negative, so by the theorem of Tonelli (1909), we can change the

order of integration:

$$\int_0^\infty \int_{-u_i-\log v}^\infty \exp(-\bar{G}v) \left((\bar{G}_i \bar{G}_j - \bar{G}_{ij} \frac{1}{v} \right) e^{u_i+u_j} v^{-\rho/(1-\rho)} e^{-(\epsilon_{it}+u_i)/(1-\rho)} d\epsilon_{it} dv$$

Evaluate the interior integral:

$$\int_0^\infty \exp(-\bar{G}v) \left((\bar{G}_i \bar{G}_j - \bar{G}_{ij} \frac{1}{v} \right) e^{u_i+u_j} v^{-\rho/(1-\rho)} [-(1-\rho) e^{-(\epsilon_{it}+u_i)/(1-\rho)}]_{-u_i-\log v}^\infty dv$$

Plug in the limits of integration,

$$(1-\rho) e^{u_i+u_j} \int_0^\infty \exp(-\bar{G}v) \left((\bar{G}_i \bar{G}_j - \bar{G}_{ij} \frac{1}{v} \right) v^{-\rho/(1-\rho)} v^{1/(1-\rho)} dv$$

Simplify,

$$(1-\rho) e^{u_i+u_j} \int_0^\infty \exp(-\bar{G}v) \left((\bar{G}_i \bar{G}_j v - \bar{G}_{ij} \right) dv$$

Evaluate the integral,

$$(1-\rho) e^{u_i+u_j} \left[-v \frac{\bar{G}_i \bar{G}_j}{\bar{G}} e^{-\bar{G}v} - \frac{\bar{G}_i \bar{G}_j}{\bar{G}^2} e^{-\bar{G}v} + \frac{\bar{G}_{ij}}{\bar{G}} e^{-\bar{G}v} \right]_0^\infty$$

This expression simplifies to:

$$(1-\rho) e^{u_i+u_j} \left(\frac{\bar{G}_i \bar{G}_j}{\bar{G}^2} - \frac{\bar{G}_{ij}}{\bar{G}} \right)$$

Recall that we have defined $\bar{G} = G(e^{u_1}, ..., e^{u_I})$, $\bar{G}_i \equiv G_i(e^{u_1}, ..., e^{u_I})$, and $\bar{G}_{ij} \equiv G_{ij}(e^{u_1}, ..., e^{u_I})$. So this completes our proof of the first expression.

For the second formulation of the proposition, recall that $p_i = e^{u_i} \frac{\bar{G}_i}{\bar{G}}$. If we factor that out,

$$(1-\rho) p_i p_j \left(1 - \frac{\bar{G}_{ij} \bar{G}}{\bar{G}_i \bar{G}_j} \right)$$

□

A.3 Proof of Corollary 1

We aim to prove:

The cross-elasticity of population to utility is given by the migration rate times

a constant.

$$\frac{\partial p_i}{\partial u_j} = -\frac{1}{1-\rho} m_{i \to j}$$

when $i \neq j$.

Proof: Recall that

$$p_i = e^{u_i} \frac{G_i(e^{u_1}, ..., e^{u_I})}{G(e^{u_1}, ..., e^{u_I})}$$

The derivative of p_i with respect to u_j is:

$$\frac{\partial p_i}{\partial u_j} = e^{u_i + u_j} \left(-\frac{G_i(e^{u_1}, ..., e^{u_I}) G_j(e^{u_1}, ..., e^{u_I})}{G(e^{u_1}, ..., e^{u_I})^2} + \frac{G_{ij}(e^{u_1}, ..., e^{u_I})}{G(e^{u_1}, ..., e^{u_I})} \right)$$

By equation (5), we can substitute $\frac{1}{1-\rho} m_{i \to j}$ for the right-hand side. □

A.4 Proof of Lemma 1

We aim to prove:

> Define Λ^x to be the sum of x i.i.d. mean-zero logistic random variables. Then in steady-state of the SPACE model,
>
> $$\lim_{\rho \to 1} \frac{m_{i \to j}^t}{m_{i \to j}^1} = \frac{\mathbb{E}[|\Lambda^{2t}|]}{\mathbb{E}[|\Lambda^2|]}$$
>
> where $\mathbb{E}[|\cdot|]$ denotes the mean absolute deviation (MAD).

Before we prove Lemma 1, we will need the following sub-lemma.

Lemma 2. *Define Υ to be the cumulative distribution function of the conditional distribution $\Delta\epsilon_n = \frac{\epsilon_{nt+1} - \epsilon_{nt}}{1-\rho}$ given ϵ_{nt}. Then,*

$$\lim_{\rho \to 1} \Upsilon(\Delta\epsilon_1, \Delta\epsilon_2, \Delta\epsilon_3, ... | \epsilon_{1t}, \epsilon_{2t}, \epsilon_{3t}, ...) = \frac{1}{1+e^{-\Delta\epsilon_1}} \frac{1}{1+e^{-\Delta\epsilon_2}} \frac{1}{1+e^{-\Delta\epsilon_3}} ...$$

Proof: The conditional distribution of ϵ_{t+1} given ϵ_t is given by

$$F(\epsilon_{t+1} | \epsilon_t) \equiv \frac{\frac{\partial^I F_2}{\prod_i \partial \epsilon_{it}}(\epsilon_{1t}, \epsilon_{2t}, ...\epsilon_{It}, \epsilon_{1t+1}, \epsilon_{2t+1}, ...\epsilon_{It+1})}{\frac{\partial^I F_2}{\prod_i \partial \epsilon_{it}}(\epsilon_{1t}, \epsilon_{2t}, ...\epsilon_{It}, \infty, \infty, ..., \infty)}$$

Recall that $F_2(\epsilon_{1t}, \epsilon_{2t}, ...\epsilon_{It}, \epsilon_{1t+1}, \epsilon_{2t+1}, ...\epsilon_{It+1}) = F(-\log H(e^{-\epsilon_{1t}}, e^{-\epsilon_{1t+1}}), ..., -\log H(e^{-\epsilon_{It}}, e^{-\epsilon_{It+1}}))$.

Define $H_i = H(e^{-\epsilon_{it}}, e^{-\epsilon_{it+1}})$ so the numerator is given by:

$$\frac{\partial^I F_2}{\prod_i \partial \epsilon_{it}} = \frac{\partial^I F}{\prod_i \partial(-\log H_i)}(-\log H_1, ..., -\log H_I) \prod_i \frac{1}{H_i} \frac{\partial H_i}{\partial e^{-\epsilon_{it}}} e^{-\epsilon_{it}}$$

by the chain rule. Recall that $H_i = (e^{-\frac{\epsilon_{it}}{1-\rho}} + e^{-\frac{\epsilon_{it+1}}{1-\rho}})^{1-\rho}$, so

$$\frac{\partial H_i}{\partial e^{-\epsilon_{it}}} = (e^{-\frac{\epsilon_{it}}{1-\rho}} + e^{-\frac{\epsilon_{it+1}}{1-\rho}})^{-\rho} e^{-\epsilon_{it}\frac{\rho}{1-\rho}}$$

Define $f = \frac{\partial^I F}{\prod_i \partial \epsilon_{it}}$. Then the numerator is given by

$$\frac{\partial^I F_2}{\prod_i \partial \epsilon_{it}}(\epsilon_{1t}, \epsilon_{2t}, ...\epsilon_{It}, \epsilon_{1t+1}, \epsilon_{2t+1}, ...\epsilon_{It+1}) = f(-\log H_1, ..., -\log H_I) \prod_i \frac{e^{-\frac{\epsilon_{it}}{1-\rho}}}{e^{-\frac{\epsilon_{it}}{1-\rho}} + e^{-\frac{\epsilon_{it+1}}{1-\rho}}}$$

For the denominator, note that $F_2(\epsilon_{1t}, \epsilon_{2t}, ...\epsilon_{It}, \infty, \infty, ..., \infty) = F(\epsilon_{1t}, \epsilon_{2t}, ...\epsilon_{It})$, so

$$\frac{\partial^I F_2}{\prod_i \partial \epsilon_{it}}(\epsilon_{1t}, \epsilon_{2t}, ...\epsilon_{It}, \infty, \infty, ..., \infty) = f(\epsilon_{1t}, ..., \epsilon_{It})$$

The whole expression is therefore:

$$\frac{f(-\log H_1, ..., -\log H_I)}{f(\epsilon_{1t}, ..., \epsilon_{It})} \prod_i \frac{e^{-\frac{\epsilon_{it}}{1-\rho}}}{e^{-\frac{\epsilon_{it}}{1-\rho}} + e^{-\frac{\epsilon_{it+1}}{1-\rho}}}$$

Recall that we defined $\Delta\epsilon_i = \frac{\epsilon_{it+1} - \epsilon_{it}}{1-\rho}$. So

$$-\log H_i = \epsilon_{it} - (1-\rho)\log(1 + e^{-\Delta\epsilon_i})$$

and

$$\frac{e^{-\frac{\epsilon_{it}}{1-\rho}}}{e^{-\frac{\epsilon_{it}}{1-\rho}} + e^{-\frac{\epsilon_{it+1}}{1-\rho}}} = \frac{1}{1 + e^{-\Delta\epsilon_i}}$$

Therefore, we can write

$$\Upsilon(\Delta\epsilon_1, \Delta\epsilon_2, \Delta\epsilon_3, ...|\epsilon_{1t}, \epsilon_{2t}, \epsilon_{3t}, ...)$$
$$= \frac{f(\epsilon_{1t} - (1-\rho)\log(1 + e^{-\Delta\epsilon_1}), ..., \epsilon_{It} - (1-\rho)\log(1 + e^{-\Delta\epsilon_I}))}{f(\epsilon_{1t}, ..., \epsilon_{It})} \prod_i \frac{1}{1 + e^{-\Delta\epsilon_i}}$$

Taking the limit as $\rho \to 1$ cancels out the f terms:

$$\lim_{\rho \to 1} \Upsilon(\Delta\epsilon_1, ...\Delta\epsilon_I | \epsilon_{1t}, ...\epsilon_{It}) = \prod_i \frac{1}{1 + e^{-\Delta\epsilon_i}}$$

□

This lemma implies that when ρ is close to 1, the changes in the ϵ's are well-approximated by independent logistic distributions.

Proof of Lemma 1: The migration probability over s periods is given by:

$$m^s_{i \to j} = \mathbb{P}(u_i + \epsilon_{in0} > \max_{k \neq i} u_k + \epsilon_{kn0} \text{ and } u_j + \epsilon_{jns} > \max_{k \neq j} u_k + \epsilon_{kns})$$

We can express this as an integral over all the ϵ.

$$m^s_{i \to j} = \int_{\epsilon_{i0} > \max_{k \neq i}\{u_k - u_i + \epsilon_{k0}\}, \quad \epsilon_{js} > \max_{k \neq j}\{u_k - u_j + \epsilon_{ks}\}} dF(\epsilon_s | \epsilon_0) dF(\epsilon_0)$$

We can rewrite the second condition in the integral using the definition of $\Delta\epsilon$. This also allows us to express $dF(\epsilon_s|\epsilon_0)$ in terms of Υ.

$$m^s_{i \to j} = \int_{\substack{\epsilon_{i0} > \max_{k \neq i}\{u_k - u_i + \epsilon_{k0}\}, \\ (1-\rho)\sum_{r=1}^s \Delta\epsilon_{jr} + \epsilon_{j0} > \max_{k \neq j}\{u_k - u_j + \sum_{r=1}^s (1-\rho)\Delta\epsilon_{kr} + \epsilon_{k0}\}}} d\Upsilon(\Delta\epsilon_s | \epsilon_{s-1}) \cdots d\Upsilon(\Delta\epsilon_2 | \epsilon_1) d\Upsilon(\Delta\epsilon_1 | \epsilon_0) dF(\epsilon_0)$$

Define $\kappa_j = \sum_{r=1}^s \Delta\epsilon_{jr}$. This is the cumulative change in ϵ_j relative to ϵ_i. Also define $d\Upsilon^s(\kappa|\epsilon_0) = \prod_{t=1}^s d\Upsilon(\Delta\epsilon_t|\epsilon_{t-1})$, the distribution of this cumulative change, given ϵ_0. Therefore, the s-year migration is given by:

$$m^s_{i \to j} = \int_{\substack{\epsilon_{i0} > \max_k\{u_k - u_i + \epsilon_{k0}\}, \\ (1-\rho)\kappa_j + \epsilon_{j0} > \max_k\{u_k - u_j + (1-\rho)\kappa_k\}}} d\Upsilon^s(\kappa|\epsilon_0) dF(\epsilon_0)$$

In order to make further progress, we first need to quantify the mass of individuals that are initially close to indifference between i and j, and whom are living in i at $t = 0$. Consider this normalized measure of how many people are almost-indifferent:

$$\begin{aligned}
\zeta_{ij} &= \lim_{\rho \to 1} \frac{1}{1-\rho} \mathbb{P}(u_i - u_j - (1-\rho) < \epsilon_{j0} - \epsilon_{i0}, u_i + \epsilon_{i0} > \max_{k \neq i} u_k + \epsilon_{k0}) \\
&= \lim_{\rho \to 1} \frac{1}{1-\rho} \int_{\substack{u_i + \epsilon_{i0} > \max_{k \neq_i} u_k + \epsilon_{k0} \\ \epsilon_{i0} - \epsilon_{j0} < u_j - u_i + (1-\rho)}} dF(\epsilon_0)
\end{aligned}$$

Intuitively, if $u_j - u_i$ increases by δ for some tiny δ, then $\delta\zeta_{ij}$ mass of people will move from i to j.

Also note that

$$\lim_{\rho\to 1}\frac{1}{1-\rho}\int_{\substack{u_i+\epsilon_{i0}>\max_{k\neq i,j}u_k+\epsilon_{k0}\\ u_j-u_i<\epsilon_{i0}-\epsilon_{j0}<u_j-u_i+(1-\rho)\\ u_k-u_i<\epsilon_{i0}-\epsilon_{k0}<u_k-u_i+A(1-\rho)}} dF(\epsilon_0) = 0 \tag{12}$$

when $j \neq k$ for any constant A. In other words, the probability of being roughly indifferent between three places vanishes more quickly that two places, as ρ tends to 1. Therefore, we will not have to keep track of the mass of people that would have switched from i to j, but k's utility increased even more.

Now consider the number of people that will switch from i to j for a given set of κ, as $\rho \to 1$:

$$\lim_{\rho\to 1}\frac{1}{1-\rho}\int_{\substack{\epsilon_{i0}>\max_{k\neq i}\{u_k-u_i+\epsilon_{k0}\},\\ (1-\rho)\kappa_j+\epsilon_{j0}>\max_{k\neq j}\{u_k-u_j+(1-\rho)\kappa_k\}}} dF(\epsilon_0)$$

Because of (12), we can drop the $\max_k$ and only consider j and i:

$$\lim_{\rho\to 1}\frac{1}{1-\rho}\int_{\substack{\epsilon_{i0}>\max_{k\neq i}\{u_k-u_i+\epsilon_{k0}\},\\ (1-\rho)\kappa_j+\epsilon_{j0}>u_i-u_j+(1-\rho)\kappa_i+\epsilon_{i0}}} dF(\epsilon_0)$$

We can change variables; define $1-\rho^* = (1-\rho)(\kappa_j - \kappa_i)$:

$$\lim_{\rho^*\to 1}\frac{\kappa_j-\kappa_i}{1-\rho^*}\int_{\substack{\epsilon_{i0}>\max_{k\neq i}\{u_k-u_i+\epsilon_{k0}\},\\ \epsilon_{j0}+(1-\rho^*)>u_i-u_j+\epsilon_{i0}}} dF(\epsilon_0)$$

Note that this is only positive if $\kappa_j > \kappa_i$; otherwise no one would move from i to j. Then simplifying using the definition of ζ_{ij}:

$$\zeta_{ij}(\kappa_j - \kappa_i)^+$$

We are now ready to consider $\lim_{\rho\to 1}\frac{m^s_{i\to j}}{1-\rho}$:

$$\lim_{\rho\to 1}\frac{m^s_{i\to j}}{1-\rho} = \lim_{\rho\to 1}\frac{1}{1-\rho}\int_{\substack{\epsilon_{i0}>\max_{k}\{u_k-u_i+\epsilon_{k0}\},\\ (1-\rho)\kappa_j+\epsilon_{j0}>\max_{k}\{u_k-u_j+(1-\rho)\kappa_k\}}} d\Upsilon^s(\kappa|\epsilon_0)dF(\epsilon_0)$$

Take the limit as $\rho \to 1$:

$$\lim_{\rho\to 1}\frac{m^s_{i\to j}}{1-\rho} = \zeta_{ij}\int_{\kappa_j-\kappa_i>0}(\kappa_j-\kappa_i)d\Upsilon^s(\kappa)$$

Note that Υ^s does not depend on ϵ_0 because of Lemma 2.

$\Upsilon^s(\kappa)$ is the cumulative distribution function of the sum of s independent logistic distributions for each i. Since logistics distributions are symmetric, we can rewrite this as

$$\lim_{\rho \to 1} \frac{m^s_{i \to j}}{1-\rho} = \zeta_{ij} \int_0^\infty \kappa d\Lambda^{2s}(\kappa)$$

where Λ^t is a sum of t i.i.d. logistic distributions.

Finally, this implies that

$$\lim_{\rho \to 1} \frac{m^s_{i \to j}}{m^1_{i \to j}} = \frac{\zeta_{ij}}{\zeta_{ij}} \frac{\int_0^\infty \kappa d\Lambda^{2s}(\kappa)}{\int_0^\infty \kappa d\Lambda^{2}(\kappa)} = \frac{\int_0^\infty \kappa d\Lambda^{2s}(\kappa)}{\int_0^\infty \kappa d\Lambda^{2}(\kappa)}$$

For a symmetric distribution $X \sim F$, such as the sum of i.i.d. logistics, $\int_0^\infty x dF(x) = \frac{1}{2}\mathbb{E}[|X|]$. Therefore,

$$\lim_{\rho \to 1} \frac{m^s_{i \to j}}{m^1_{i \to j}} = \frac{\mathbb{E}[|\Lambda^{2s}|]}{\mathbb{E}[|\Lambda^{2}|]}$$

□

A.5 Proof of Proposition 3

We aim to prove:

> In the steady-state of the SPACE model, as $\rho \to 1$, the ratio of t-year migration to 1-year migration for any state-pair is bounded below by $\sqrt{t}$ and bounded above by $\sqrt{\pi/3}\sqrt{t}$:
>
> $$\sqrt{t} \leq \lim_{\rho \to 1} \frac{m^t_{i \to j}}{m^1_{i \to j}} \leq \sqrt{\frac{\pi}{3}}\sqrt{t}$$

Proof: We proceed in two steps. First we show that the expression of interest is increasing in s. Then we find the limit as $s \to \infty$, which is an upper bound. Since it is trivially equal to 1 for $s = 2$, that establishes the lower bound.

The expression of interest is weakly increasing

First, we would like to show that

$$\frac{1}{\sqrt{s}} \int_{-\infty}^0 \Lambda^s(\kappa) d\kappa$$

is weakly increasing. If we integrate by parts,

$$\frac{1}{\sqrt{s}}\int_{-\infty}^{0}\Lambda^s(\kappa)d\kappa = \frac{1}{\sqrt{s}}\int_{-\infty}^{0}-\kappa\lambda^s(\kappa)d\kappa = \frac{1}{2\sqrt{s}}\mathbb{E}[|X^s|]$$

where the second equality holds because the distribution is symmetric, and we define λ^s to be the pdf, and X^s to be a random variable with distribution Λ^s. The absolute value function admits the following Fourier integral representation:

$$|x| = \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{1-e^{-iux}}{u^2}du$$

Taking expectations,

$$\mathbb{E}[|X|] = \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{1-\mathbb{E}[e^{-iuX}]}{u^2}du$$

$\mathbb{E}[e^{-iuX}]$ is the characteristic function, which is known for convolutions of logistic distributions. In particular, the characteristic function of a standard logistic distribution is:

$$\phi(u) = \frac{\pi u}{\sinh(\pi u)}$$

So the convolution of s logistic distributions is:

$$\phi^s(u) = \left(\frac{\pi u}{\sinh(\pi u)}\right)^s$$

Plugging this into the Fourier transformation of the absolute value,

$$\frac{1}{\sqrt{s}}\mathbb{E}[|X^s|] = \frac{1}{\pi}\frac{1}{\sqrt{s}}\int_{-\infty}^{\infty}\frac{1-\left(\frac{\pi u}{\sinh(\pi u)}\right)^s}{u^2}du$$

A simple u-substitution of $u' = u\sqrt{s}$:

$$\frac{1}{\sqrt{s}}\mathbb{E}[|X^s|] = \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{1-\left(\frac{\pi u/\sqrt{s}}{\sinh(\pi u/\sqrt{s})}\right)^s}{u^2}du$$

And because sinh is an odd function, the integrand is an even function. So we can simplify:

$$\frac{1}{\sqrt{s}}\mathbb{E}[|X^s|] = \frac{2}{\pi}\int_{0}^{\infty}\frac{1-\left(\frac{\pi u/\sqrt{s}}{\sinh(\pi u/\sqrt{s})}\right)^s}{u^2}du$$

Now we want to prove that this is weakly increasing in s. A sufficient condition is to

prove that

$$\left(\frac{\pi u/\sqrt{s}}{\sinh(\pi u/\sqrt{s})}\right)^s$$

is weakly decreasing in s for all u.

We will consider the log of the expression, which is weakly decreasing if and only if the expression itself is weakly decreasing.

$$f(s) = -s\left(\log\left(\frac{\sinh(\pi u/\sqrt{s})}{\pi u\sqrt{s}}\right)\right)$$

Euler proved that $\sinh(z) = z\prod_{x=1}^{\infty}(1 + z^2/(\pi^2 x^2))$ (see e.g. Bair, Błaszczyk, Ely, Henry, Kanovei, Katz, Katz, Kutateladze, McGaffey, Schaps, Sherry and Shnider, 2013).

Plugging that in,

$$f(s) = -s\sum_{x=1}^{\infty}\log\left(1 + \frac{u^2}{sx^2}\right)$$

Even though we only care about discrete s, this equation is well-defined for continuous s, and if it is weakly decreasing for all s that are continuous, then it is also weakly decreasing for all discrete s. So we can show this is weakly decreasing by taking the derivative with respect to s and showing it's non-positive.

$$f'(s) = -\sum_{x=1}^{\infty}\left(\log\left(1 + \frac{u^2}{sx^2}\right) - \frac{u^2}{sx^2 + u^2}\right)$$

To see the term inside the sum is non-negative, define $t = u^2/(sx^2)$, and note that t is non-negative for positive values of s. The term we are considering becomes $\log(1+t) - t/(t+1)$. At $t = 0$, this term is 0. Its derivative is $1/(1 + t) - 1/(1 + t)^2$, which is positive for all $t > 0$. So the term is always non-negative.

Therefore $f'(s)$ is always non-positive. Therefore, $\left(\frac{\pi u/\sqrt{s}}{\sinh(\pi u/\sqrt{s})}\right)^s$ is weakly decreasing in s for all u. Therefore, $\frac{1}{\sqrt{s}}\int_{-\infty}^{0}\Lambda^s(\kappa)d\kappa$ is weakly increasing in s

The expression of interest is bounded above by $\sqrt{\pi/3}$

Given that $\frac{1}{\sqrt{s}}\int_{-\infty}^{0}\Lambda^s(\kappa)d\kappa$ is increasing in s, we next focus on its value at $s = 2$ and as $s \to \infty$.

First, consider its limiting behavior as $s \to \infty$. Here, we can apply the central limit theorem, which tells us that the sum of many logistic variables, divided by the square root of s approaches a normal variable with standard deviation equal to that of the standard

logistic, known to be $\frac{\pi}{\sqrt{3}}$. Therefore, we consider

$$\int_{-\infty}^{0} \Phi\left(\frac{x\sqrt{3}}{\pi}\right) dx$$

where Φ is the standard normal distribution CDF. Integrating by parts,

$$u = \Phi\left(\frac{x\sqrt{3}}{\pi}\right), \quad v = x, \quad du = \sqrt{\frac{3}{2\pi^3}}\exp(-\frac{3x^2}{2\pi^2})dx, \quad dv = dx$$

So,

$$\begin{aligned}
\int_{-\infty}^{0} \Phi\left(\frac{x\sqrt{3}}{\pi}\right) dx &= \left[\Phi\left(\frac{x\sqrt{3}}{\pi}\right) x\right]_{-\infty}^{0} - \int_{-\infty}^{0} x\sqrt{\frac{3}{2\pi^3}}\exp(-\frac{3x^2}{2\pi^2})dx \\
&= \sqrt{\frac{\pi}{6}}\left[\exp\left(\frac{-3x^2}{2\pi^2}\right)\right]_{-\infty}^{0} \\
&= \sqrt{\frac{\pi}{6}}
\end{aligned}$$

We are also interested in $s = 2$, which we can calculate via integration of the logistic probability density function.

$$\begin{aligned}
\frac{1}{\sqrt{2}}\int_{-\infty}^{0} \Lambda^2(\kappa)d\kappa &= \frac{1}{\sqrt{2}}\int_{-\infty}^{0}\int_{-\infty}^{\infty}\int_{-\infty}^{\kappa-x} \frac{e^{-x}}{(1+e^{-x})^2}\frac{e^{-y}}{(1+e^{-y})^2}dydxd\kappa \\
&= \frac{1}{\sqrt{2}}\int_{-\infty}^{0}\int_{-\infty}^{\infty} \frac{e^{-x}}{(1+e^{-x})^2}\frac{1}{1+e^{-(\kappa-x)}}dxd\kappa \\
&= \frac{1}{\sqrt{2}}\int_{-\infty}^{0}\left[-\frac{e^{\kappa}(-(e^x+1)\log(e^{\kappa}+e^x)+e^{\kappa}+(e^x+1)\log(e^x+1)-1)}{(e^{\kappa}-1)^2(e^x+1)}\right]_{-\infty}^{\infty} d\kappa \\
&= \frac{1}{\sqrt{2}}\int_{-\infty}^{0} \frac{e^{\kappa}(-\kappa+e^{\kappa}-1)}{(e^{\kappa}-1)^2}d\kappa \\
&= \frac{1}{\sqrt{2}}\left[\frac{\kappa}{e^{\kappa}-1}+\kappa\right]_{-\infty}^{0} \\
&= \frac{1}{\sqrt{2}}
\end{aligned}$$

So the fraction

$$\lim_{s\to\infty}\sqrt{\frac{2}{s}}\frac{\int_{-\infty}^{0}\Lambda^s(\kappa)d\kappa}{\int_{-\infty}^{0}\Lambda^2(\kappa)d\kappa} = \frac{\sqrt{\pi/6}}{\sqrt{1/2}} = \sqrt{\frac{\pi}{3}}$$

In words, this means that the ratio we are interested will converge to $\sqrt{\pi/3}$ as s approaches ∞.

The expression of interest is bounded below by 1

We also know the ratio for $s = 2$ because it is trivial:

$$\sqrt{\frac{2}{2}}\frac{\int_{-\infty}^{0}\Lambda^2(\kappa)d\kappa}{\int_{-\infty}^{0}\Lambda^2(\kappa)d\kappa} = 1$$

Completing the proof

And since $\frac{1}{\sqrt{s}}\int_{-\infty}^{0}\Lambda^s(\kappa)d\kappa$ is increasing in s it follows that

$$1 \leq \sqrt{\frac{2}{s}}\frac{\int_{-\infty}^{0}\Lambda^s(\kappa)d\kappa}{\int_{-\infty}^{0}\Lambda^2(\kappa)d\kappa} \leq \sqrt{\frac{\pi}{3}}$$

for all $s \geq 2$.

Finally, we substitute in the expression from Lemma 1:

$$1 \leq \frac{1}{\sqrt{s}}\frac{m_{i\to j}^{s}}{m_{i\to j}^{1}} \leq \sqrt{\frac{\pi}{3}}$$

□

A.6 Proof of Proposition 4

We aim to prove:

> In the long-run steady state of a moving cost model, if migration is symmetric, in the sense that $\lim_{\Delta\to\infty} m_{i\to j}/m_{j\to i} = 1$, then
>
> $$\lim_{\Delta\to\infty}\frac{\partial \log p_i}{\partial v_j} = \begin{cases} -2p_j & \text{if } i \neq j \\ 2 - 2p_j & \text{if } i = j \end{cases} \tag{13}$$

Proof: The migration rate from i to j is given by

$$\frac{m_{i\to j}}{p_i} = \frac{\exp(v_j - \delta_{ij})}{\sum_k \exp(v_k - \delta_{ik})}$$

In steady-state, it must be the case that

$$p_j = \sum_k p_k \frac{m_{k\to j}}{p_k} = \sum_k p_k \frac{\exp(v_j - \delta_{kj})}{\sum_\ell \exp(v_\ell - \delta_{k\ell})}$$

for all i. Implicitly differentiate with respect to v_i:

$$\frac{\partial p_j}{\partial v_i} = \sum_k \left[\frac{\partial p_k}{\partial v_i} \frac{\exp(v_j - \delta_{kj})}{\sum_\ell \exp(v_\ell - \delta_{k\ell})} - p_k \frac{\exp(v_j - \delta_{kj}) \exp(v_i - \delta_{ki})}{(\sum_\ell \exp(v_\ell - \delta_{k\ell}))^2} \right]$$

$$\frac{\partial p_j}{\partial v_i} \left(1 - \frac{\exp(v_j)}{\sum_\ell \exp(v_\ell - \delta_{j\ell})} \right) = \sum_{k\neq j} \left[\frac{\partial p_k}{\partial v_i} \frac{\exp(v_j - \delta_{kj})}{\sum_\ell \exp(v_\ell - \delta_{k\ell})} \right] - \sum_k \left[p_k \frac{\exp(v_j - \delta_{kj}) \exp(v_i - \delta_{ki})}{(\sum_\ell \exp(v_\ell - \delta_{k\ell}))^2} \right]$$

$$\frac{\partial p_j}{\partial v_i} \left(\frac{\sum_{\ell\neq j} \exp(v_\ell - \delta_{j\ell})}{\sum_\ell \exp(v_\ell - \delta_{j\ell})} \right) = \sum_{k\neq j} \left[\frac{\partial p_k}{\partial v_i} \frac{\exp(v_j - \delta_{kj})}{\sum_\ell \exp(v_\ell - \delta_{k\ell})} \right] - \sum_k \left[p_k \frac{\exp(v_j - \delta_{kj}) \exp(v_i - \delta_{ki})}{(\sum_\ell \exp(v_\ell - \delta_{k\ell}))^2} \right]$$

Divide both sides by $\exp(-\Delta)$:

$$\frac{\partial p_j}{\partial v_i} \left(\frac{\sum_{\ell\neq j} \exp(v_\ell - \delta_{j\ell})}{\sum_\ell \exp(v_\ell - \delta_{j\ell}) \exp(-\Delta)} \right) = \sum_{k\neq j} \left[\frac{\partial p_k}{\partial v_i} \frac{\exp(v_j - \delta_{kj})}{\sum_\ell \exp(v_\ell - \delta_{k\ell}) \exp(-\Delta)} \right] - \sum_k \left[p_k \frac{\exp(v_j - \delta_{kj}) \exp(v_i - \delta_{ki})}{(\sum_\ell \exp(v_\ell - \delta_{k\ell}))^2 \exp(-\Delta)} \right]$$

Now take the limit as $\Delta \to \infty$:

$$\frac{\partial p_j}{\partial v_i} \left(\frac{\sum_{\ell\neq j} \exp(v_\ell - \delta'_{j\ell})}{\exp(v_j)} \right) = \sum_{k\neq j} \left[\frac{\partial p_k}{\partial v_i} \frac{\exp(v_j - \delta'_{kj})}{\exp(v_k)} \right] - \left[p_j \frac{\exp(v_i - \delta'_{ji})}{\exp(v_j)} \right] - \left[p_i \frac{\exp(v_j - \delta'_{ij})}{\exp(v_i)} \right] \tag{14}$$

This is a system of $I - 1$ equations.

For the last equation, consider

$$\frac{\partial p_i}{\partial v_i} = \sum_k \left[\frac{\partial p_k}{\partial v_i} \frac{\exp(v_i - \delta_{ki})}{\sum_\ell \exp(v_\ell - \delta_{k\ell})} - p_k \frac{\exp(v_i - \delta_{ki}) \exp(v_i - \delta_{ki})}{(\sum_\ell \exp(v_\ell - \delta_{k\ell}))^2} + p_k \frac{\exp(v_i - \delta_{ki})}{\sum_\ell \exp(v_\ell - \delta_{k\ell})} \right]$$

Do the same transformations:

$$\frac{\partial p_i}{\partial v_i} \left(\frac{\sum_{\ell \neq i} \exp(v_\ell - \delta_{i\ell})}{\sum_\ell \exp(v_\ell - \delta_{i\ell}) \exp(-\Delta)} \right) = \sum_{k \neq i} \left[\frac{\partial p_k}{\partial v_i} \frac{\exp(v_i - \delta_{ki})}{\sum_\ell \exp(v_\ell - \delta_{k\ell}) \exp(-\Delta)} \right] + \sum_k \left[p_k \frac{\exp(v_i - \delta_{ki})}{\sum_\ell \exp(v_\ell - \delta_{k\ell})} \left(\frac{\sum_{\ell \neq i} \exp(v_\ell - \delta_{k\ell})}{\sum_\ell \exp(v_\ell - \delta_{k\ell}) \exp(-\Delta)} \right) \right]$$

Then the limit as $\Delta \to \infty$ is:

$$\frac{\partial p_i}{\partial v_i} \left(\frac{\sum_{\ell \neq i} \exp(v_\ell - \delta'_{i\ell})}{\exp(v_i)} \right) = \sum_{k \neq i} \left[\frac{\partial p_k}{\partial v_i} \frac{\exp(v_i - \delta'_{ki})}{\exp(v_k)} \right] + \sum_{k \neq i} p_k \frac{\exp(v_i - \delta'_{ki})}{\exp(v_k)} + p_i \sum_{k \neq i} \frac{\exp(v_k - \delta'_{ik})}{\exp(v_i)} \tag{15}$$

We will guess the following solution to (14) and (15):

$$\frac{\partial p_i}{\partial v_i} = 2p_i - 2p_i^2$$
$$\frac{\partial p_j}{\partial v_i} = -2p_i p_j$$

Equation (14) becomes:

$$(-2p_i p_j) \left(\frac{\sum_{\ell \neq j} \exp(v_\ell - \delta'_{j\ell})}{\exp(v_j)} \right) = \sum_{k \neq j,i} \left[-2p_i p_k \frac{\exp(v_j - \delta'_{kj})}{\exp(v_k)} \right] + (2p_i - 2p_i^2) \frac{\exp(v_j - \delta'_{ij})}{\exp(v_i)} - p_j \frac{\exp(v_i - \delta'_{ji})}{\exp(v_j)} - p_i \frac{\exp(v_j - \delta'_{ij})}{\exp(v_i)}$$

This becomes:

$$(-2p_i p_j) \left(\frac{\sum_{\ell \neq j} \exp(v_\ell - \delta'_{j\ell})}{\exp(v_j)} \right) = \sum_{k \neq j} \left[-2p_i p_k \frac{\exp(v_j - \delta'_{kj})}{\exp(v_k)} \right] + p_i \frac{\exp(v_j - \delta'_{ij})}{\exp(v_i)} - p_j \frac{\exp(v_i - \delta'_{ji})}{\exp(v_j)}$$

In steady-state, we know that $p_j \frac{\sum_{\ell \neq j} \exp(v_\ell - \delta_{j\ell})}{\sum_\ell \exp(v_\ell - \delta_{j\ell})} = \sum_{k \neq j} p_k \frac{\exp(v_j - \delta_{kj})}{\sum_\ell \exp(v_\ell - \delta_{k\ell})}$. Multiplying

by $\exp(\Delta)$ and taking the limit as $\Delta \to \infty$, $p_j \frac{\sum_{\ell \neq j} \exp(v_\ell - \delta'_{j\ell})}{\exp(v_j)} = \sum_{k \neq j} p_k \frac{\exp(v_j - \delta'_{kj})}{\exp(v_k)}$. So the first two terms cancel out.

Similarly, we assumed that $\lim_{\Delta \to \infty} m_{i \to j}/m_{j \to i} = 1$, which implies that $p_i \frac{\exp(v_j - \delta'_{ij})}{\exp(v_i)} = p_j \frac{\exp(v_i - \delta'_{ji})}{\exp(v_j)}$. So the second two terms also cancel out. Therefore, equation (14) is satisfied.

Equation (15) becomes:

$$(2p_i - 2p_i^2)\left(\frac{\sum_{\ell \neq i} \exp(v_\ell - \delta'_{i\ell})}{\exp(v_i)}\right) = \sum_{k \neq i}\left[(-2p_i p_k)\frac{\exp(v_i - \delta'_{ki})}{\exp(v_k)}\right] + \sum_{k \neq i} p_k \frac{\exp(v_i - \delta'_{ki})}{\exp(v_k)} + p_i \sum_{k \neq i} \frac{\exp(v_k - \delta'_{ik})}{\exp(v_i)}$$

All terms cancel out because, as we showed, $p_j \frac{\sum_{\ell \neq j} \exp(v_\ell - \delta'_{k\ell})}{\exp(v_j)} = \sum_{k \neq j} p_k \frac{\exp(v_j - \delta'_{kj})}{\exp(v_j)}$. So equation (15) is also satisfied, meaning that

$$\frac{\partial p_i}{\partial v_i} = 2p_i - 2p_i^2$$
$$\frac{\partial p_j}{\partial v_i} = -2p_i p_j$$

is a solution to the problem. To express the equations as semi-elasticities, divide both sides by p_j.

□

B Calibration

In this section, we go into details on our preferred calibration. As stated in the main text, we wish to calibrate a model where

$$G(x_1, ..., x_I) = \sum_{q=1}^{Q} \left(\sum_{i \in \mathcal{N}_q} (w_{iq} x_i)^{\frac{1}{1-\gamma}}\right)^{1-\gamma}, \tag{16}$$

where $1 - \gamma = \frac{1-\rho}{1-\tilde{\rho}}$, where $\tilde{\rho}$ is a constant, and $\rho \to 1$. As we discussed in the text, rather than calibrating w_{iq}, we will calibrate

$$\tilde{w}_{iq} \equiv (w_{iq} e^{u_i})^{\frac{1}{1-\gamma}} \left(\sum_{k \in \mathcal{N}_q} (w_{kq} e^{u_k})^{\frac{1}{1-\gamma}}\right)^{-\gamma}$$

Under these definitions, we can derive the two expressions for populations and migration as functions of these $\tilde{w}$, equations (8) and (9).

Let us start with equation (8). Recall that $p_i = e^{u_i}\frac{G_i}{G}$, and that we choose w_{ij} such that $G = 1$. G_i is given by

$$G_i(x_1, ..., x_I) = \sum_{q=1}^{Q} \left(\sum_{k \in \mathcal{N}_q} (w_{kq}x_k)^{\frac{1}{1-\gamma}} \right)^{-\gamma} (w_{iq}x_i)^{\frac{1}{1-\gamma}} x_i^{-1}$$

Plugging in the e^{v_i},

$$e^{u_i}G_i(e^{u_1}, ..., e^{u_I}) = \sum_{q=1}^{Q} \left(\sum_{k \in \mathcal{N}_q} (w_{kq}e^{u_k})^{\frac{1}{1-\gamma}} \right)^{-\gamma} (w_{iq}e^{u_i})^{\frac{1}{1-\gamma}}$$

The term inside the summation is $\tilde{w}_{iq}$. So

$$p_i = \sum_{q=1}^{Q} \tilde{w}_{iq} \tag{17}$$

which is equation (8) (note that $w_{iq} = 0$ if $i \notin \mathcal{N}_q$).

Now, let us derive equation (9). Recall that $m_{i\to j} = (1-\rho)e^{u_i}e^{u_j}\left(\frac{G_iG_j}{G^2} - \frac{G_{ij}}{G}\right)$. Note that this can be rewritten $m_{i\to j} = (1-\rho)p_ip_j - (1-\rho)e^{u_i}e^{u_j}\frac{G_{ij}}{G}$. The limit of the first term as $\rho \to 1$ is 0. For the second term, G_{ij} is given by

$$G_{ij}(x_1, ...x_I) = -\frac{\gamma}{1-\gamma} \sum_{q=1}^{Q} \left(\sum_{k \in \mathcal{N}_q} (w_{kq}x_k)^{\frac{1}{1-\gamma}} \right)^{-\gamma-1} (w_{iq}x_i)^{\frac{1}{1-\gamma}} x_i^{-1} (w_{jq}x_j)^{\frac{1}{1-\gamma}} x_j^{-1}$$

So

$$m_{i\to j} = -(1-\rho)\frac{\gamma}{1-\gamma} \sum_{q=1}^{Q} \left(\sum_{k \in \mathcal{N}_q} (w_{kq}x_k)^{\frac{1}{1-\gamma}} \right)^{-\gamma-1} (w_{iq}e^{u_i})^{\frac{1}{1-\gamma}} (w_{jq}e^{u_j})^{\frac{1}{1-\gamma}}$$

Note that $\sum_{i\in\mathcal{N}_q}\tilde{w}_{iq} = \left(\sum_{i\in\mathcal{N}_q}(w_{iq}x_i)^{\frac{1}{1-\gamma}}\right)^{1-\gamma}$. So this whole term can be rewritten

$$m_{i\to j} = -(1-\rho)\frac{\gamma}{1-\gamma}\sum_{q=1}^{Q}\left(\sum_{k\in\mathcal{N}_q}(w_{kq}x_k)^{\frac{1}{1-\gamma}}\right)^{\gamma-1} \times\left(\sum_{k\in\mathcal{N}_q}(w_{kq}x_k)^{\frac{1}{1-\gamma}}\right)^{-\gamma}(w_{iq}e^{u_i})^{\frac{1}{1-\gamma}}\left(\sum_{k\in\mathcal{N}_q}(w_{kq}x_k)^{\frac{1}{1-\gamma}}\right)^{-\gamma}(w_{jq}e^{u_j})^{\frac{1}{1-\gamma}}$$

which simplifies to

$$m_{i\to j} = -(1-\rho)\frac{\gamma}{1-\gamma}\sum_{q=1}^{Q}\tilde{w}_{iq}\tilde{w}_{jq}(\sum_{k\in\mathcal{N}_q}\tilde{w}_{kq})^{-1}$$

Now, taking the limit as $\rho\to 1$ and remembering that $1-\gamma = \frac{1-\rho}{1-\tilde{\rho}}$,

$$m_{i\to j} = (1-\tilde{\rho})\sum_{q=1}^{Q}\frac{\tilde{w}_{iq}\tilde{w}_{jq}}{\sum_{k\in\mathcal{N}_q}\tilde{w}_{kq}}$$

which is equation (9).

When we consider the two-location nests, it will be helpful to write this equation as:

$$\frac{1-\tilde{\rho}}{m_{i\to j}} = \frac{1}{\tilde{w}_{ij}} + \frac{1}{\tilde{w}_{ji}} \tag{18}$$

Recall that we defined $\tilde{w}_{ij}\equiv\tilde{w}_{iq}$ when $q=\{i,j\}$.

The goal of the two-location nest calibration is to maximize $\tilde{\rho}$ subject to (17) and (18) and the constraints that the $\tilde{w}$ must be positive. To do this, we will set up a Lagrangian and set its derivatives equal to zero. Call the Lagrange multipliers on (17), λ_i and the ones on (18), ν_{ij}. We will initially solve the Lagrangian without constraining the $\tilde{w}_{ij}$ to be non-negative and will later verify the solution satisfies this constraint. Then the Lagrangian is given by

$$\tilde{\rho} - \lambda_i\left(p_i - \tilde{w}_i - \sum_{j\neq i}\tilde{w}_{ij}\right) - \nu_{ij}\left(\frac{1-\tilde{\rho}}{m_{i\to j}} - \frac{1}{\tilde{w}_{ij}} - \frac{1}{\tilde{w}_{ji}}\right)$$

The first-order conditions with respect to $\tilde{w}_{ij}$ and $\tilde{w}_{ji}$, which are necessary for it to be

a maximum are

$$\lambda_i = \frac{1}{\tilde{w}_{ij}^2}\nu_{ij}$$

and

$$\lambda_j = \frac{1}{\tilde{w}_{ji}^2}\nu_{ij}$$

So $\frac{\tilde{w}_{ij}}{\tilde{w}_{ji}} = \sqrt{\frac{\lambda_j}{\lambda_i}}$. We can use (18), along with this new formulation to solve for the $\tilde{w}_{ij}$ in terms of these Lagrange multipliers:

$$\tilde{w}_{ij} = \frac{1}{1-\tilde{\rho}} m_{i\to j}\left(1 + \sqrt{\frac{\lambda_j}{\lambda_i}}\right) \tag{19}$$

Plugging (19) into (17) and multiplying by $\sqrt{\lambda_i}$,

$$(1-\tilde{\rho})\sqrt{\lambda_i} = \sum_{j\neq i} \frac{m_{i\to j}}{p_i}\left(\sqrt{\lambda_i} + \sqrt{\lambda_j}\right) \tag{20}$$

This is a necessary condition for $\tilde{\rho}$ to be maximized.

Define the $I \times I$ matrix M as

$$M_{ij} = \begin{cases} \frac{m_{i\to j}}{p_i} & \text{if } i \neq j \\ \frac{m_i}{p_i} & \text{if } i = j \end{cases}$$

(recall that $m_i = \sum_{j\neq i} m_{i\to j}$). Then we can rewrite equation (20) as

$$(1-\tilde{\rho})\ell = M\ell$$

where ℓ is an $I \times 1$ vector and $\ell_i = \sqrt{\lambda_i}$.[44]

This equation is satisfied for any eigenvector ℓ of matrix M, and corresponding eigenvalue $(1-\tilde{\rho})$. However, we also require that both ℓ and $(1-\tilde{\rho})$ are positive. When $m_{i\to j} > 0$, the matrix has all positive values, so there exists a unique real eigenvalue corresponding to an eigenvector with all positive entries by the Perron (1907)-Frobenius (1912) theorem. That eigenvalue is our calibration for $(1-\tilde{\rho})$, and the corresponding eigenvector tells us the $\sqrt{\lambda_i}$. We can use that eigenvector to calculate $\tilde{w}_{ij}$ from (19), and, if we are interested, we can calculate w_{ij} from $w_{ij} = e^{-u_i}\tilde{w}_{ij}^{1-\gamma}(\tilde{w}_{ij} + \tilde{w}_{ji})^{\gamma}$.

[44]The matrix M should be based on steady-state migration. So for $m_{i\to j}$, we recommend using the logarithmic average of $m_{i\to j}$ and $m_{j\to i}$ based on our analysis in Appendix B.4.

B.1 Exact Hat Algebra Derivation

Consider the calibration from Section 3.2. Here, we show the derivation for the exact hat algebra. We would like to express $\hat{p}_i \equiv p_i'/p_i$ as a function of $\hat{u}_i \equiv \exp((u_i' - u_i)/(1-\rho))$ in the limit as $\rho \to 1$.

For now, assume that G does not change with u_i in the limit as $\rho \to 1$. We can verify this at the end by checking whether $\sum_i p_i \hat{p}_i = 1$. Under this assumption, $p_i = G_i(u_1, ...u_I)e^{u_i}$ and $p_i' = G_i(u_1', ..., u_I')e^{u_i'}$. With the functional form we assumed for G, consider p_i'

$$p_i' = \sum_{q=1}^{Q} \left(\sum_{k \in \mathcal{N}_q} (w_{kq} e^{u_k'})^{\frac{1}{1-\gamma}} \right)^{-\gamma} (w_{iq} e^{u_i'})^{\frac{1}{1-\gamma}}$$

This can be rewritten

$$p_i' = \sum_{q=1}^{Q} \left(\sum_{k \in \mathcal{N}_q} (w_{kq} e^{u_k})^{\frac{1}{1-\gamma}} \hat{u}_k^{1-\tilde{\rho}} \right)^{-\gamma} (w_{iq} e^{u_i})^{\frac{1}{1-\gamma}} \hat{u}_i^{1-\tilde{\rho}}$$

Recall that $w_{iq} e^{u_i} = \tilde{w}_{iq}^{1-\gamma} \left(\sum_{k \in \mathcal{N}_q} \tilde{w}_{kq} \right)^{\gamma}$.

$$p_i' = \sum_{q=1}^{Q} \left(\sum_{k \in \mathcal{N}_q} \left(\tilde{w}_{kq}^{1-\gamma} \left(\sum_{\ell \in \mathcal{N}_q} \tilde{w}_{\ell q} \right)^{\gamma} \right)^{\frac{1}{1-\gamma}} \hat{u}_k^{1-\tilde{\rho}} \right)^{-\gamma} \tilde{w}_{iq} \left(\sum_{\ell \in \mathcal{N}_q} \tilde{w}_{\ell q} \right)^{\frac{\gamma}{1-\gamma}} \hat{u}_i^{1-\tilde{\rho}}$$

Simplifying,

$$p_i' = \sum_{q=1}^{Q} \left(\sum_{k \in \mathcal{N}_q} \tilde{w}_{kq} \hat{u}_k^{1-\tilde{\rho}} \right)^{-\gamma} \tilde{w}_{iq} \left(\sum_{\ell \in \mathcal{N}_q} \tilde{w}_{\ell q} \right)^{\gamma} \hat{u}_i^{1-\tilde{\rho}}$$

Recall that as $\rho \to 1$, $\gamma \to 1$ as well, so

$$p_i' = \sum_{q=1}^{Q} \tilde{w}_{iq} \frac{\sum_{\ell \in \mathcal{N}_q} \tilde{w}_{\ell q}}{\sum_{k \in \mathcal{N}_q} \tilde{w}_{kq} \hat{u}_k^{1-\tilde{\rho}}} \hat{u}_i^{1-\tilde{\rho}}$$

To express it as $\hat{p}_i$, divide both sides by p_i:

$$\hat{p}_i = \sum_{q=1}^{Q} \frac{\tilde{w}_{iq}}{p_i} \frac{\sum_{\ell \in \mathcal{N}_q} \tilde{w}_{\ell q}}{\sum_{k \in \mathcal{N}_q} \tilde{w}_{kq} \hat{u}_k^{1-\tilde{\rho}}} \hat{u}_i^{1-\tilde{\rho}}$$

This is the expression we wished to show.

It remains to show that $\sum_i p_i \hat{p}_i = 1$, to check our assumption from the beginning. This is equal to

$$\sum_i p_i \hat{p}_i = \sum_i \sum_{q=1}^{Q} \tilde{w}_{iq} \frac{\sum_{\ell \in \mathcal{N}_q} \tilde{w}_{\ell q}}{\sum_{k \in \mathcal{N}_q} \tilde{w}_{kq} \hat{u}_k^{1-\tilde{\rho}}} \hat{u}_i^{1-\tilde{\rho}}$$

If we interchange the sums,

$$\sum_i p_i \hat{p}_i = \sum_{q=1}^{Q} \frac{\sum_i \tilde{w}_{iq} \hat{u}_i^{1-\tilde{\rho}}}{\sum_{k \in \mathcal{N}_q} \tilde{w}_{kq} \hat{u}_k^{1-\tilde{\rho}}} \sum_{\ell \in \mathcal{N}_q} \tilde{w}_{\ell q}$$

Simplifying,

$$\sum_i p_i \hat{p}_i = \sum_{q=1}^{Q} \sum_{\ell \in \mathcal{N}_q} \tilde{w}_{\ell q} = 1$$

B.2 Alternative calibration with nest size greater than 2

In this section, we present a calibration in which some nests include more than 2 states. The strategy to build nests will be to choose a set of states, and allocate a fixed percentage of the population of each state to the nest. We will choose that fixed proportion to be the maximum that we can that does not over-predict migration between any two states within the nest. After we have done this for a number of nests, we will exactly match any remaining migration by allocating the remaining population to singleton or pairwise nests that will exactly match migration. For now, consider a fixed $\tilde{\rho}$.

For the first nest, choose $\mathcal{N}_1 = \{1, ..., I\}$, the entire set of locations. Define

$$\alpha_N = \frac{1}{1-\tilde{\rho}} \min_{i,j} \frac{m_{i \to j}}{p_i p_j}$$

and pick

$$\tilde{w}_{i1} = \alpha_N p_i$$

This means that an α_N share of the population chooses this nest, and this nest generates

$$m^1_{i \to j} = (1-\tilde{\rho}) \alpha_N p_i p_j$$

migration between i and j. α_N was chosen to the largest possible without exceeding the predicted migration for any pair of locations.

Next, each Census region will have its own nest. First define residual migration:

$$\bar{m}^1_{i\to j} = m_{i\to j} - m^1_{i\to j}$$

Then label the regions $r = 2, 3, 4, 5$, and for each region R_r, define a nest $\mathcal{N}_r = \{i : i \in R_r\}$, and choose population shares:

$$\alpha_r = \frac{1}{1-\tilde{\rho}} p_r \min_{i,j\in R} \frac{\bar{m}^1_{i\to j}}{p_i p_j}$$

where $p_r = \sum_{i\in R_r} p_i$. For each i in region R_r, pick

$$\tilde{w}_{ir} = \begin{cases} \alpha_r p_i & \text{if } i \in R_r \\ 0 & \text{otherwise} \end{cases}$$

Each nest generates migration:

$$m^r_{i\to j} = (1-\tilde{\rho})\alpha_r \frac{p_i p_j}{p_r}$$

if $i, j \in R_r$, and 0 otherwise. Iteratively define residual migration:

$$\bar{m}^s_{i\to j} = \bar{m}^{s-1}_{i\to j} - m^s_{i\to j}$$

for $s \geq 2$. Next, repeat the same steps for the nine divisions, labeled $d = 6, ..., 14$. Here, $\alpha_d = \frac{1}{1-\tilde{\rho}} p_d \min_{i,j\in d} \frac{\bar{m}^5_{i\to j}}{p_i p_j}$ and pick $\tilde{w}_d = \alpha_d p_i$ if $i \in D_d$ and 0 otherwise. These nests generate migration $m^d_{i\to j} = (1-\tilde{\rho})\alpha_d \frac{p_i p_j}{p_r}$ if $i, j \in D_d$, and 0 otherwise.

Repeat the steps one more time with the following 15 subgroups, labeled 15 to 29: (15) California, Washington, and Oregon; (16) Kansas, Oklahoma, Missouri, and Nebraska; (17) Indiana, Michigan, and Illinois; (18) Louisiana, Texas, Arkansas, and New Mexico; (19) Montana, South Dakota, and North Dakota; (20) Colorado, Nevada, and Arizona; (21) New Hampshire, Maine, and Vermont; (22) Florida, Mississippi, and Alabama; (23) Iowa, Minnesota, and Wisconsin; (24) Connecticut, New York, Rhode Island, and Massachusetts; (25) Georgia, South Carolina, Tennessee, and North Carolina; (26) West Virginia, Ohio, and Kentucky; (27) Delaware, New Jersey, and Pennsylvania; (28) Wyoming, Idaho, and Utah; (29) D.C., Maryland, and Virginia.

At this point, there is still residual migration $\bar{m}^{29}_{i\to j}$. To exactly match this migration,

create pairwise nests $\mathcal{N}_x = \{i, j\}$, for $x = 30, ..., X$, and choose

$$\alpha_x = \frac{1}{1-\tilde{\rho}}(p_i + p_j)\frac{\bar{m}^{29}_{i\to j}}{p_i p_j}$$

and pick

$$\tilde{w}_{i,x} = \begin{cases} \alpha_x p_i & \text{if } i \in \mathcal{N}_x \\ 0 & \text{otherwise} \end{cases}$$

This fully accounts for all internal migration. The remaining population is allocated to nests with only the one location in it.

This will be guaranteed to match migration patterns exactly. However, it requires that

$$\sum_{q=1}^{X} \tilde{w}_{iq} \leq p_i$$

for all i so that we do not violate the nonnegativity constraint on the population in the one-location nests. Because $\frac{1}{1-\tilde{\rho}}$ scales all the α, this problem is a constraint on the $\tilde{\rho}$ we can use. In practice, it requires that

$$\tilde{\rho} \leq .841$$

This is not that much less than the $\tilde{\rho}$ from the main calibration, which was 0.892.

Under this calibration, about 1.8 percent of people are in a nest with all locations, 2.4 percent of people are in a region-nest, 1.5 percent of people are in a division-nest, 2.1 percent are in nests defined by the groups above, and 48.2 percent of are in a pairwise nest. So a non-zero share of people will end up living in more than two locations over many years.

B.3 Exact Hat Algebra for Moving Cost Model

In the moving cost model, population is given by:

$$p'_{it} = \sum_j p'_{jt-1} \frac{\exp(v'_i - \delta_{ji})}{\sum_k \exp(v'_k - \delta_{jk})}$$

Expressing this in terms of $\hat{p}_{it} \equiv p'_{it}/p_{it}$:

$$\hat{p}_{it}p_{it} = \sum_j \hat{p}_{jt-1}p_{jt-1}\hat{v}_i \frac{\exp(v_i - \delta_{ji})}{\sum_k \hat{v}_k \exp(v_k - \delta_{jk})}$$

Note that $\frac{m_{j\to k}}{m_{j\to i}} = \frac{\exp(v_k - \delta_{jk})}{\exp(v_i - \delta_{ji})}$, so

$$\hat{p}_{it} = \hat{v}_i \sum_j \hat{p}_{jt-1}p_{jt-1} \frac{1}{p_{it}} \frac{m_{j\to i}}{\sum_k \hat{v}_k m_{j\to k}}$$

Rearranging,

$$\hat{p}_{it} = \hat{v}_i \sum_j \frac{\frac{m_{j\to i}}{p_i}\hat{p}_{jt-1}}{\sum_k \frac{m_{j\to k}}{p_j}\hat{v}_k}$$

This is the equation in the text.

B.4 Non-steady-state migration

In our calibration, we can still solve for migration even if the model is not in steady-state. In particular, in a two-region model where the ϵ's are independent across space and the correlation across time is governed by $\tilde{\rho}$, migration from i to j is defined as:

$$m_{i\to j} = P(\epsilon_{i1} + u_{i1} > \epsilon_{j1} + u_{j1}, \epsilon_{i2} + u_{i2} < \epsilon_{j2} + u_{j2})$$

where the distribution of the ϵ_{it} are given by:

$$F(\epsilon_{i1}, \epsilon_{j1}, \epsilon_{i2}, \epsilon_{j2}) = \exp\left(-\left(e^{-\frac{\epsilon_{i1}}{1-\tilde{\rho}}} + e^{-\frac{\epsilon_{i2}}{1-\tilde{\rho}}}\right)^{1-\tilde{\rho}} - \left(e^{-\frac{\epsilon_{j1}}{1-\tilde{\rho}}} + e^{-\frac{\epsilon_{j2}}{1-\tilde{\rho}}}\right)^{1-\tilde{\rho}}\right)$$

Let us define $u_1 = u_{i1} - u_{j1}$ and $u_2 = u_{i2} - u_{j2}$. Without loss of generality, assume $u_2 > u_1$. This implies that $m_{i\to j} > m_{j\to i}$ Then the probability is given by

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{\partial^2 F}{\partial \epsilon_{i1}\epsilon_{j2}}(\epsilon_{i1}, \epsilon_{i1} + u_1, \epsilon_{j2} - u_2, \epsilon_{j2}) d\epsilon_{i1} d\epsilon_{j2}$$

Writing out the derivative,

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\exp\left(-\left(e^{-\frac{\epsilon_{i1}}{1-\tilde{\rho}}}+e^{-\frac{\epsilon_{j2}-u_2}{1-\tilde{\rho}}}\right)^{1-\tilde{\rho}}-\left(e^{-\frac{\epsilon_{i1}+u_1}{1-\tilde{\rho}}}+e^{-\frac{\epsilon_{j2}}{1-\tilde{\rho}}}\right)^{1-\tilde{\rho}}\right)$$
$$\times\left(e^{-\frac{\epsilon_{i1}}{1-\tilde{\rho}}}+e^{-\frac{\epsilon_{j2}-u_2}{1-\tilde{\rho}}}\right)^{-\tilde{\rho}}e^{-\frac{\epsilon_{i1}}{1-\tilde{\rho}}}\left(e^{-\frac{\epsilon_{i1}+u_1}{1-\tilde{\rho}}}+e^{-\frac{\epsilon_{j2}}{1-\tilde{\rho}}}\right)^{-\tilde{\rho}}e^{-\frac{\epsilon_{j2}}{1-\tilde{\rho}}}d\epsilon_{i1}d\epsilon_{j2}$$

Consider the following two-dimensional u-substitution:

$$v=-(e^{-\frac{\epsilon_{i1}}{1-\tilde{\rho}}}+e^{-\frac{\epsilon_{j2}-u_2}{1-\tilde{\rho}}})^{1-\tilde{\rho}}$$
$$w=-(e^{-\frac{\epsilon_{i1}+u_1}{1-\tilde{\rho}}}+e^{-\frac{\epsilon_{j2}}{1-\tilde{\rho}}})^{1-\tilde{\rho}}$$
$$dvdw=\left(1-\exp\left(\frac{u_1-u_2}{1-\tilde{\rho}}\right)\right)\left(e^{-\frac{\epsilon_{i1}}{1-\tilde{\rho}}}+e^{-\frac{\epsilon_{j2}-u_2}{1-\tilde{\rho}}}\right)^{-\tilde{\rho}}e^{-\frac{\epsilon_{i1}}{-\tilde{\rho}}}\left(e^{-\frac{\epsilon_{i1}+u_1}{1-\tilde{\rho}}}+e^{-\frac{\epsilon_{j2}}{1-\tilde{\rho}}}\right)^{1-\tilde{\rho}}e^{-\frac{\epsilon_{j1}}{1-\tilde{\rho}}}d\epsilon_{i1}d\epsilon_{j2}$$

where the term multiplying $d\epsilon_{i1}d\epsilon_{j2}$ is the absolute value of the Jacobian. The integral becomes

$$\frac{1}{1-\exp\left(\frac{u_1-u_2}{1-\tilde{\rho}}\right)}\int_{-\infty}^{0}\int_{\exp(u_1)w}^{\exp(u_2)w}\exp(v+w)dvdw$$

Taking the interior integral,

$$\frac{1}{1-\exp(\frac{u_1-u_2}{1-\tilde{\rho}})}\int_{-\infty}^{0}\exp((e^{u_2}+1)w)-\exp((e^{u_1}+1)w)dw$$

Taking the second integral,

$$\frac{1}{1-\exp(\frac{u_1-u_2}{1-\tilde{\rho}})}\left(\frac{1}{1+e^{u_2}}-\frac{1}{1+e^{u_1}}\right)$$

In the calibration, this expression represents the share of $\tilde{w}_{ij}+\tilde{w}_{ji}$ people that move from i to j in the period when the relative utility changes from u_1 to u_2. The expression for the other direction can be found by switching $-u_1$ for u_1 and $-u_2$ for u_2. Note that the limit as u_1 approaches u_2 is the expression for the migration in steady-state.

Now that we have solved for non-steady state migration, we can use this expression to solve for what steady-state migration would have been when we observe $m_{i\to j}$ and $m_{j\to i}$ in the data. We are interested in the amount of migration there would have been had we been in a steady-state at u_2, since it is steady-state migration that pins down the parameters of $\tilde{w}_{ij}$ and, more importantly, the cross-elasticities of population. This is

given by

$$m^{ss}_{i\to j} = (1-\tilde{\rho})\frac{e^{u_2}}{(1+e^{u_2})^2}$$

Given

$$m_{i\to j} = \frac{1}{1-\exp(\frac{u_1-u_2}{1-\tilde{\rho}})}\left(\frac{1}{1+e^{u_2}} - \frac{1}{1+e^{u_1}}\right)$$
$$m_{j\to i} = \frac{1}{1-\exp(\frac{u_2-u_1}{1-\tilde{\rho}})}\left(\frac{1}{1+e^{-u_2}} - \frac{1}{1+e^{-u_1}}\right)$$

which we observe in the data, and a calibrated $1-\tilde{\rho}$, we can solve for u_2, and therefore find $m^{ss}_{i\to j}$.

Note that $\frac{m_{i\to j}}{m_{j\to i}} = \exp\left(\frac{u_2-u_1}{1-\tilde{\rho}}\right)$. So $e^{u_1} = e^{u_2}(m_{j\to i}/m_{i\to j})^{1-\tilde{\rho}}$. Then

$$m_{i\to j} = \frac{1}{1-m_{j\to i}/m_{i\to j}}\left(\frac{1}{1+e^{u_2}} - \frac{1}{1+e^{u_2}(m_{j\to i}/m_{i\to j})^{1-\tilde{\rho}}}\right)$$

Simplifying,

$$m_{i\to j} - m_{j\to i} = \frac{e^{u_2}((m_{j\to i}/m_{i\to j})^{1-\tilde{\rho}} - 1)}{(1+e^{u_2})(1+e^{u_2}(m_{j\to i}/m_{i\to j})^{1-\tilde{\rho}})}$$

At this point, we could solve a quadratic equation for e^{u_2} and plug that into $m^{ss}_{i\to j}$.[45] However, there is not much economic intuition in that solution, so here we also consider an approximation when $\tilde{\rho} \approx 1$. In particular, we can rearrange the equation to be

$$\frac{m_{i\to j} - m_{j\to i}}{\left(\frac{(m_{j\to i}/m_{i\to j})^{1-\tilde{\rho}}-1}{1-\tilde{\rho}}\right)} = m^{ss}_{i\to j}\frac{1+e^{u_2}}{1+e^{u_2}(m_{j\to i}/m_{i\to j})^{1-\tilde{\rho}}}$$

If we take the limit as $\tilde{\rho} \to 1$, then

$$\lim_{\tilde{\rho}\to 1}\frac{1}{m^{ss}_{i\to j}}\frac{m_{i\to j} - m_{j\to i}}{\log(m_{i\to j}/m_{j\to i})} = 1$$

So when $\tilde{\rho} \approx 1$, the logarithmic mean of the migrations is a good approximation of the

[45] In particular, $e^{u_2} = \frac{-b-\sqrt{b^2-4ac}}{2a}$ where

$$a = (m_{i\to j} - m_{j\to i})(m_{j\to i}/m_{i\to j})^{1-\tilde{\rho}}$$
$$b = (m_{i\to j} - m_{j\to i}) + (m_{i\to j} - m_{j\to i})(m_{j\to i}/m_{i\to j})^{1-\tilde{\rho}} - (m_{j\to i}/m_{i\to j})^{1-\tilde{\rho}} + 1$$
$$c = m_{i\to j} - m_{j\to i}$$

and $m^{ss}_{i\to j} = (1-\tilde{\rho})\frac{e^{u_2}}{(1+e^{u_2})^2}$. Note that the other root of the quadratic equation gives you e^{-u_1}, which without loss of generality, we had assumed to be larger.

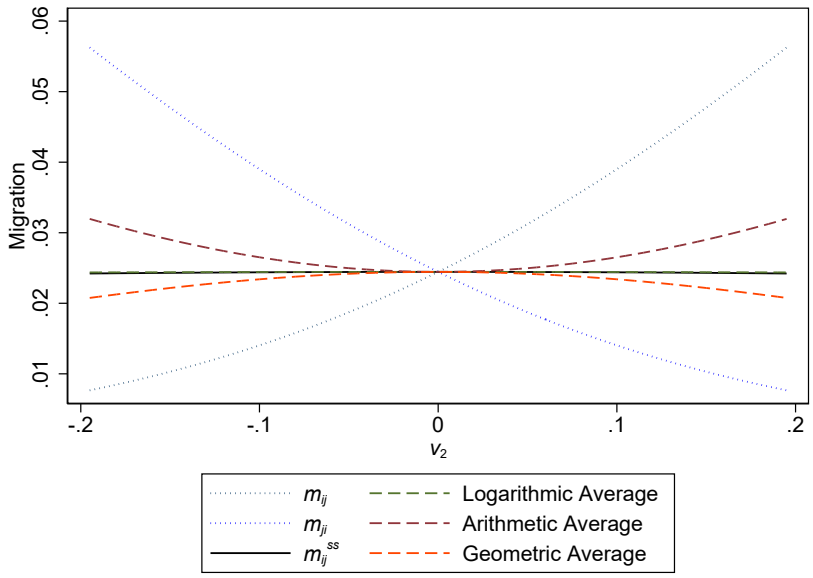


Figure A1: Bilateral migration outside of steady-state. In this figure, we plot the $m_{i\to j}$ migration rate, the $m_{j\to i}$ migration rate when $v_1 = 0$ and v_2 is the value on the x-axis. We also plot $m^{ss}_{i\to j}$, corresponding to $v_1 = v_2$. We then plot several averages of $m_{i\to j}$ and $m_{j\to i}$ to demonstrate that the logarithmic average gives the best approximation.

steady-state migration.

In Figure A1, we plot $m^{ss}_{i\to j}$ along with $m_{i\to j}$ and $m_{j\to i}$, and several candidate approximations. To create this figure, we calculate $m_{i\to j}$, $m_{j\to i}$ and $m^{ss}_{i\to j}$ using the formulae above, assuming that $u_1 = 0$ and that u_2 ranges from -0.25 to 0.25. This implies that the period 1 populations are 0.5 in both i and j, and each population ranges from about 0.45 to 0.55 in period 2. We then plot the logarithmic, the arithmetic, and the geometric average of $m_{i\to j}$ and $m_{j\to i}$ for each value of u_2, to be able to compare how good of an approximation each one is. The logarithmic approximation is visually indistinguishable from the true $m^{ss}_{i\to j}$. The other two are good approximations if $u_2 \approx 0$, but there are sizable gaps for larger u_2.

Of course, in the data, it is typically true that $m_{i\to j} \approx m_{j\to i}$ (see Figure A7a). So the precise method for taking the average is unlikely to make a large difference.

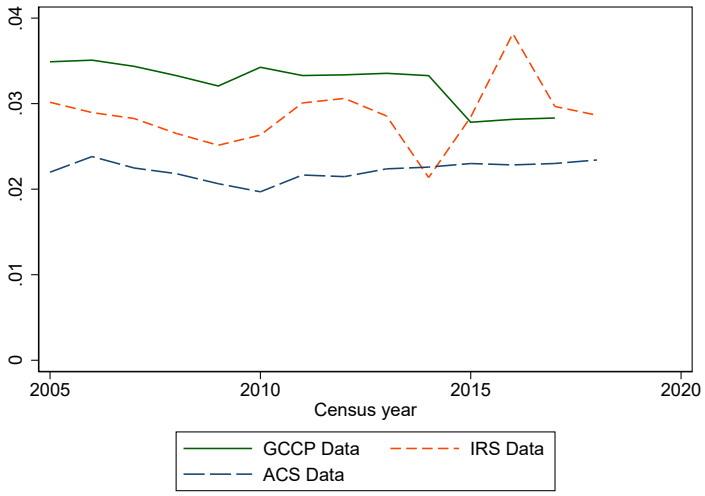


Figure A2: Comparison of 1-year interstate migration rates in GCCP, IRS, and ACS data.

C Appendix tables and figures

C.1 Migration in the GCCP

The data are consistent with two well-known facts that are often the motivation of standard moving cost models. The first fact is that interstate migration is rare. In Figure A2, we show that the GCCP records just over 3 percent of Americans moving between states in any given year. This is slightly higher than the IRS data or ACS data, which is also shown in the figure. The vast majority of Americans do not move between states in any given year.[46]

The second fact is that migration follows a gravity pattern, meaning that the amount of migration between two states is increasing in each state's population, and decreasing in the distance between them. In Table A1, we show the results from a pseudo-Poisson maximum likelihood regression in which we regress migration on log distance and the log populations of the origin and destination states (Silva and Tenreyro, 2006; Correia, Guimaraes and Zylkin, 2019):

$$\log m_{i \to j} = \beta \log \text{distance}_{ij} + \alpha \log p_i + \gamma \log p_j + \epsilon_{ij} \tag{21}$$

where p_i is the population of i and $m_{i \to j}$ is the 1-year migration from i to j. To calculate distances between states, we use the population-weighted centroid of each state from Mis-

[46]The PSID data cover an earlier time period, 1969-1997, so are not included in Figure A2. The average migration rate during that time period is 2.69 percent.

Table A1: Gravity Equations

	(1) Migration (IRS)	(2) Migration (GCCP)	(3) Migration (IRS)	(4) Migration (GCCP)
Log Distance	-0.662***	-0.666***	-1.065***	-1.037***
	(0.0597)	(0.0547)	(0.0697)	(0.0685)
Log Origin Population	0.886***	0.910***		
	(0.0837)	(0.0813)		
Log Destination Population	0.814***	0.882***		
	(0.101)	(0.0829)		
Observations	2550	2550	2550	2550
Origin and Destination FEs			Yes	Yes

Standard errors are two-way clustered by origin and destination states

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

souri Data Census Center (2000) and calculate distances using the formula from Vincenty (1975), which accounts for the fact that the Earth is a spheroid, not a perfect sphere. As expected, we find that β is negative and both α and γ are positive. More importantly, the coefficients for the IRS data and the GCCP data are similar. We also show a specification with origin state and destination state fixed effects, and the coefficients on distance are again similar across the datasets.

C.2 Square root fact table

In Table A2, we show the point estimates, as well as standard errors and number of data points from Figure 1(a). To account for the fact that some of the migration rates are from individuals that show up in more than one year in the data (and sometimes include overlapping years), the standard errors are clustered by individual. Nonetheless, they are three or four orders of magnitude smaller than the estimates themselves.

Table A2: t-year migration rate

Panel A: 1-7 years							
	(1) 1 year	(2) 2 years	(3) 3 years	(4) 4 years	(5) 5 years	(6) 6 years	(7) 7 years
Migration Rate	0.0326	0.0504	0.0650	0.0776	0.0879	0.0974	0.106
	(0.0000452)	(0.0000680)	(0.0000880)	(0.000106)	(0.000122)	(0.000137)	(0.000151)
Observations	39268194	34734474	30778061	27209466	23923278	20861568	18054919
Panel B: 8-14 years							
	8 years	9 years	10 years	11 years	12 years	13 years	14 years
Migration Rate	0.114	0.121	0.128	0.135	0.142	0.149	0.156
	(0.000164)	(0.000178)	(0.000192)	(0.000206)	(0.000221)	(0.000238)	(0.000260)
Observations	15532462	13043817	10644689	8335294	6114681	3987744	1939471

Standard errors are clustered by individual.

C.3 Distribution of the square root fact

In this subsection, we examine how the square root fact is distributed across state-pairs and age cohorts to see if it holds generally or is specific to our aggregation. To create Figure A3a, we calculate $m^t_{i\to j}$, the migration rate from state i to state j at time t, for all pairs of states i and j, using data from the GCCP. Because migration volumes vary significantly between different pairs (e.g., migration between California and Texas is much higher than between Wyoming and Vermont), we apply a normalization to ensure comparability. Specifically, we consider $c_{ij}m^t_{i\to j}$, where c_{ij} is given by

$$c_{ij} = \frac{\sum_{x=1}^{14}\sqrt{x}}{\sum_{t=1}^{14} m^t_{i\to j}}$$

which scales the migration rates so they are comparable for states with varying amounts of migration. We then plot the 5th, 25th, 50th (median), 75th, and 95th percentiles, along with the mean, of the normalized migration rates $c_{ij}m^t_{i\to j}$ across all state-pairs. These percentiles and means are weighted by the total migration volume $\sum_{t=1}^{14} m^t_{i\to j}$ for each pair. The results indicate that the square root fact holds consistently across state-pairs, with relatively little deviation.

A similar analysis is presented in Figure A3b, where we examine the square root fact across different ages rather than state-pairs. Age is calculated at the end of the migration period. So if someone is 47 in 2007, and we wish to calculate the three-year migration rate for 47 year olds, then we would use the 2004-2007 period to calculate the value. We use the same normalization and weighting scheme as we did for state-pairs. Again, the distribution closely aligns with the square root relationship, though there is slightly more variation compared to the state-pair analysis.

A similar analysis is also presented in Figure A3c, where we examine the square root fact across different birth years. Birth year is calculated by subtracting age from the current year. Again, we use the same normalization and weighting scheme. The distribution closely aligns with the square root relationship.

In Figure A3d, we examine the square root fact by the starting year. For example, the line labeled 2010 plots the migration from 2010-2011, 2010-2012, 2010-2013, etc. Since there are only 14 potential starting years, we graph all 14, without any normalization. There is very little variation across years.

In Figure A4a, we split states into whether they are close together or far apart, based on the migration-weighted median distance between states (about 1000 km). We normalize

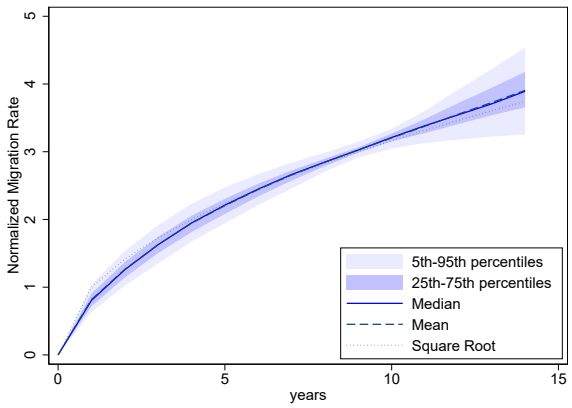
the curves as we did for state-pairs. There is no particular difference in the square root pattern between close states and far away states.

In Figure A4b, we split our sample to be between 2004-2011, and 2011-2018, and plot the t-year migration rates over those horizons. There is no meaningful difference between the two time periods. Note that for these, we can only look for the square root pattern up to the 7-year migration rate, since that is the length of half our sample.

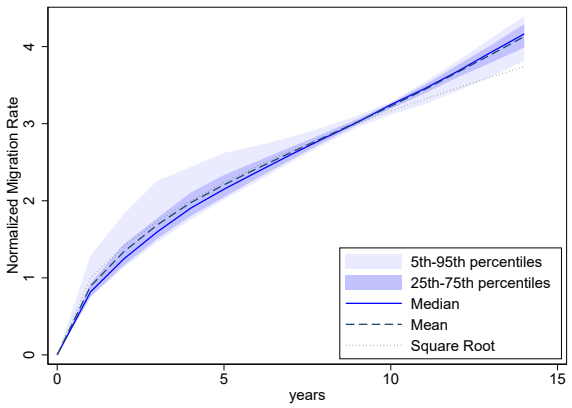
C.4 Age and the square root fact

A reader might worry that migration in the GCCP is inaccurately measured for young people, and that the mismeasurement could be driving the square root fact. For example, college students might not be measured accurately. As we show in Figure A5a, younger people are much more likely to move, and so if mismeasurement were an issue for this group, one could imagine that it would drive the square root fact.

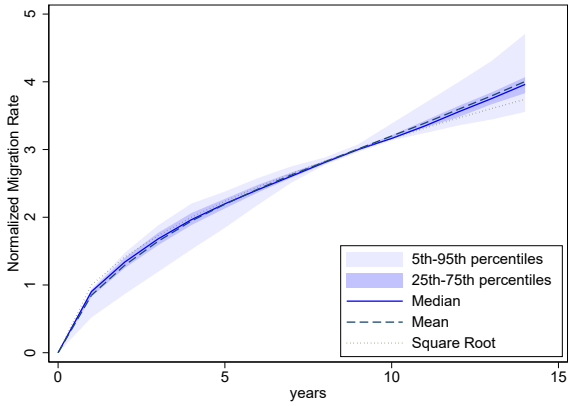
To address this concern, we split the sample at the median age, 45, and graph the square root fact for people above or below that age. If, over the course of the t years, they turn 45, they are not included in either calculation. So the 5-year migration rate for young people would include a person who started at age 35 and ended at age 40, and for old people would include a person who started at age 50 and ended at 44, but someone who started at 42 and ended at 47 would not be included in the calculation. The sample split is shown in Figure A5. While the migration rates for young people are uniformly higher, a square root line is a good fit for both.



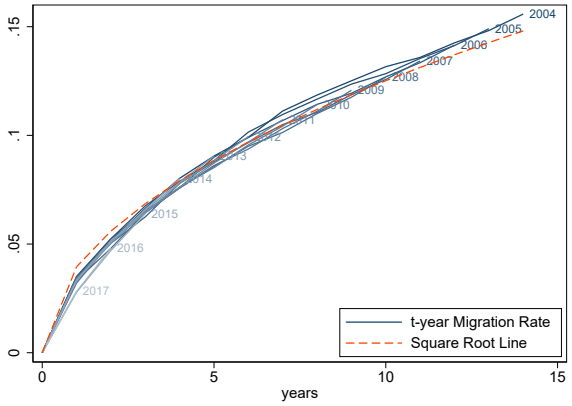
(a) Square Root Fact across state-pairs



(b) Square Root Fact across ages

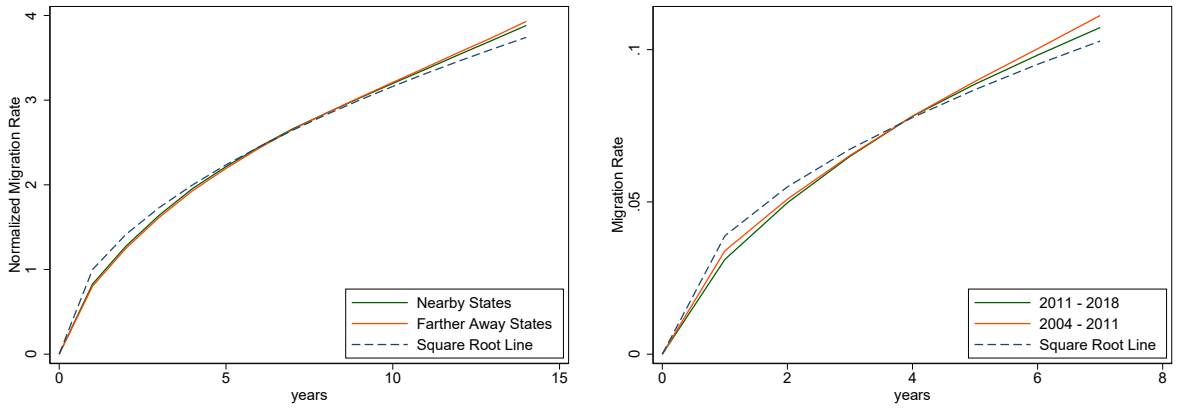


(c) Square Root Fact across birth-year cohorts



(d) Square Root Fact for each starting year

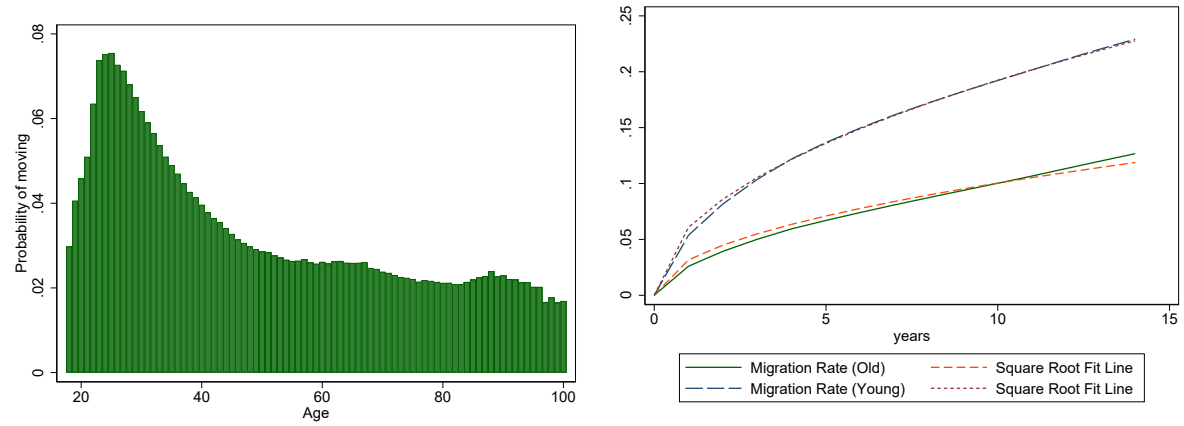
Figure A3: Distribution of the square root fact. For panel (a), for each state-pair, we calculate the share of people that move from state i to state j over t years. We then normalize these shares such that the sum of the shares over t equals $\sum_{x=1}^{14} \sqrt{s}$. This figure shows the t-by-t distribution over the (i, j) pairs, weighted by the number of migrants from i to j. For panel (b), for each age, we calculate the share of people that move from state i to state j over t years. We then normalize these shares such that the sum of the shares over t equals $\sum_{x=1}^{14} \sqrt{s}$. This figure shows the t-by-t distribution over the age cohorts, weighted by the size of the cohort. For panel (c), for each birth-year, we calculate the share of people that move from state i to state j over t years. We then normalize these shares such that the sum of the shares over t equals $\sum_{x=1}^{14} \sqrt{s}$. This figure shows the t-by-t distribution over the age cohorts, weighted by the size of the cohort. For panel (d), For each starting year, we calculate the share of people that move from state i to state j from the starting year to the starting year plus t.



(a) Square Root Fact for nearby and far away states.

(b) Square Root Fact over different time periods.

Figure A4: The Square Root Fact for various subsamples of the data. In panel (a), we plot the t-year migration rate for nearby and far-away states, splitting the number of states at the migration-weighted median. The lines are normalized as described in the text, as the nearby states have higher overall levels of migration. A square root line is provided for comparison. In panel (b), we split the sample in time, and plot the t-year migration rate using only data from 2004-2011, and separately, only data from 2011-2018.



(a) Moving probability by age in the GCCP

(b) Square root fact for young and old subsamples

Figure A5: Robustness to considering whether age explains the square root fact. In panel (a), we plot the migration probability by age. We censor the graph at 18 and 100, and plot the one-year migration rate by each age. In panel (b), we split the sample by age, using the median age of 45, and graph the t-year migration rate for each subsample. Both rates fit the square root line well.

C.5 Simulations of the Moving Cost Model

In this Appendix, we simulat the moving cost model in order to verify that the relationship between the t-year migration rate and t is approximately linear.

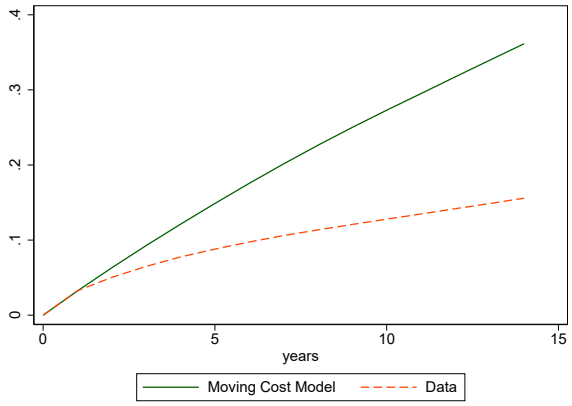
First, using the standard moving cost model, we simulate the t-year migration rate for one million agents over 15 years. In this model, migration probabilities depend only on the current location. Using the observed migration probabilities from the GCCP in 2004-2005, we then simulate the location that the agent would live in over the next 15 years, and use that simulation to calculate the t-year migration rate. Figure A6a shows the results of this simulation. The relationship between the t-year migration rate and t looks linear and does not provide a close fit to the data.

In simulations, it is also possible to calibrate migration costs that depend on variables that might be assumed could break the linear relationship. For example, many models include as a state variable whether the person moved in the previous year and, if so, from where (e.g. Kennan and Walker, 2011). To capture this in the simulation of the moving cost model, we take observed migration probabilities from 2005-2006 that depend on the location in 2004 and 2005, and then simulate future locations, based on the previous two locations, for 15 total years. We again simulate one million agents. Mechanically, this simulation will generate a realistic amount of return migration. We show the t-year migration rate in Figure A6b. While this additional state variable adds a kink at $t = 1$ and matches the data almost exactly in both $t = 1$ and $t = 2$, the relationship becomes fairly linear again for larger t's.

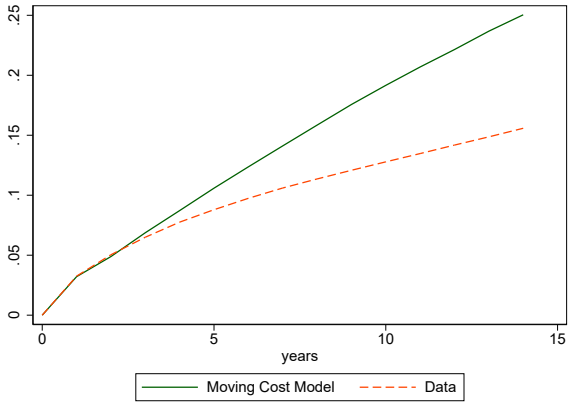
Similarly, one might think that adding age as a state variable could help match the square root fact since migration rates decline over the life-cycle (through the lens of the standard model, moving costs increase in age). Using separate moving probabilities for every age and current location, we again simulate one million agents. The t-year migration rate can be seen in Figure A6c, and it is again approximately linear.

Finally, we also check whether adding the state of birth is helpful to match the curvature of the t-year migration rate, since moving back home may be a high-enough probability event to make Proposition 1 a bad approximation.[47] For this calculation, we use the American Community Survey (ACS) data, which records the state of birth, as well as interstate migration, to calibrate the moving probabilities for the simulation. The migra-

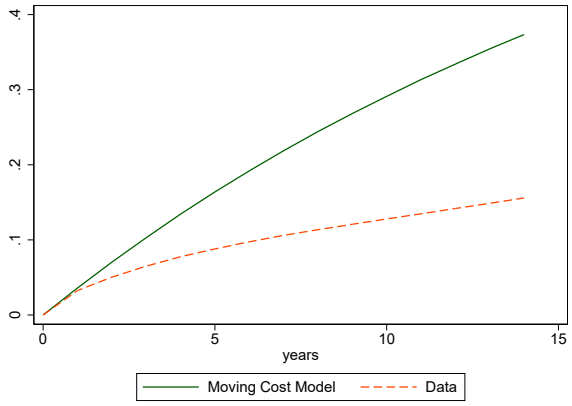
[47] The reader might wonder if other demographics would matter. For example, college educated workers are known to move more (Molloy, Smith and Wozniak, 2011). However, for immutable characteristics, the aggregate migration rate would just be the average of the different groups, so if the groups have linear t-year migration rates, then the aggregate would also be linear. So as long as migration costs are high for every group, then the t-year migration rate will still be linear.



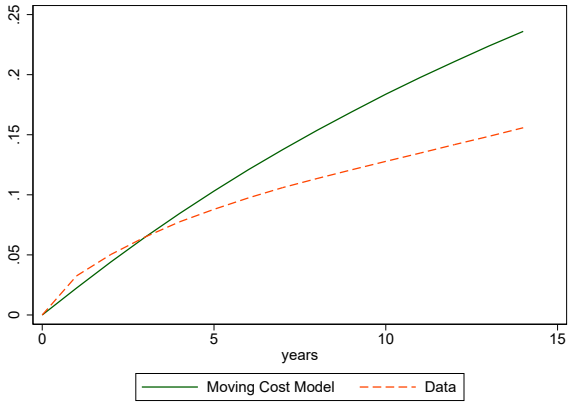
(a) Baseline Moving Cost Model



(b) Previous Year's Location State Variable



(c) Age State Variable



(d) Birthplace State Variable

Figure A6: Do existing moving cost models match the square root fact? Each panel compares the t-year migration rate in simulations of the moving cost model described in Section 2.3 to the data. In Panel (a), migration costs vary by origin-destination pair. In Panel (b), migration costs are allowed to vary by the interaction of location in the year prior, origin, and destination (i.e. someone that had lived in state X then moved to state Y will have different moving costs than someone that lived in state Y for two years). In Panel (c), migration costs vary by the interaction of age, origin, and destination. And in Panel (d), migration costs vary by the interaction of birthplace, origin, and destination. The data are from the GCCP, and for panels (a)-(c), the model is fit on the GCCP data. Panel (d) is fit using ACS data, which has a lower 1-year migration rate, and so the data and model do not match even at $t = 1$.

tion rate in the ACS is modestly lower, so the 1-year migration rate will not be as well matched as in the other simulations. Again, we simulate one million observations, and present the t-year migration rate in Figure A6d. It does not match the $\sqrt{t}$ relationship.

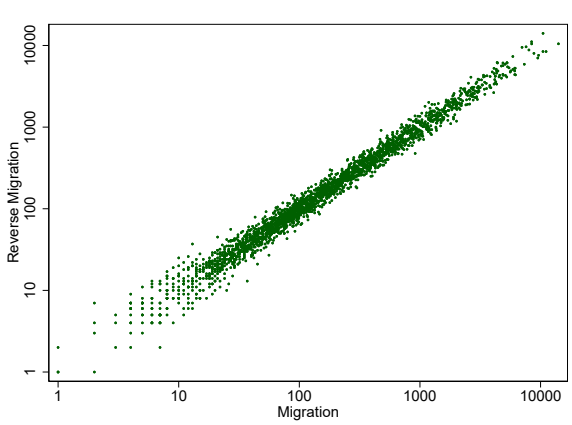
C.6 Correlation of migration with opposite-direction migration

In this appendix, we show evidence of the high degree of symmetry in the migration data. In particular, we are interested in the relationship between $m_{i \to j}$ and $m_{j \to i}$. For the data, we use the total one-year migration from 2004-2018 in the GCCP; aggregating over years reduces the measurement error due to sampling.

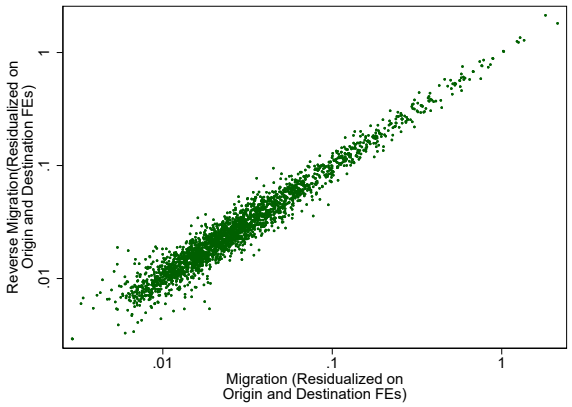
In Figure A7a, we show a scatter plot of the raw data, using a log-log scale. The data is clustered near the 45 degree line, indicating a strong empirical relationship between migration from i to j and migration from j to i. In fact, the correlation of $\log m_{i \to j}$ and $\log m_{j \to i}$ is 0.99.

Some of this correlation is likely due to the fact that some states are larger than others, and therefore have more in- and out-migration. In Panel (b), we perform a pseudo-Poisson maximum likelihood regression of the migration data on origin and destination fixed effects. We calculate residuals by dividing the data by the predicted value. We then plot the residuals against one another. Again, the data is close to the 45 degree line.

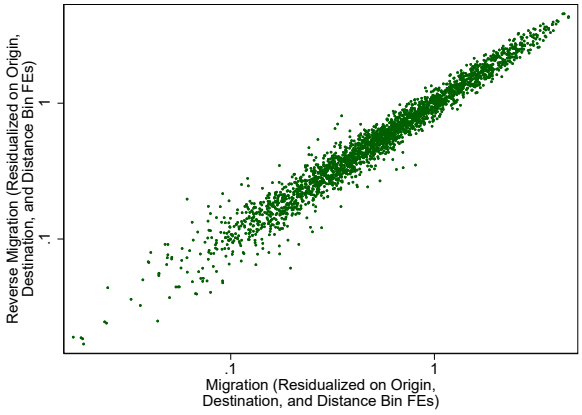
Finally, to see if this is just a function of distance, we split the data into 100 bins based on the distance between the two states (roughly 25 state-pairs per bin). We then include indicators for the bin in our pseudo-Poisson maximum likelihood regression, and plot the residuals against each other. Again, the data is close to the 45 degree line. The correlation of the logs of the residuals is now 0.92. Even conditional on origin, destination, and distance, there is a strong relationship between migration from i to j and migration from j to i.



(a) Raw data



(b) Residualized on origin and destination fixed effects



(c) Residualized on origin, destination, and distance fixed effects

Figure A7: The relationship between migration from i to j and migration from j to i. Each figure plots the total one-year migration from i to j in the GCCP, from 2004-2018. In panels (b) and (c), the migration is first regressed on a set of fixed effects using psuedo-Poisson regression, and the data is divided by the predicted value from the regression. For panel (c), state-pairs are split into 100 bins based on the distance from each other, and an indicator variable for the bin is used in the regression. Each graph uses a log scale for both axes.

C.7 The square root fact outside of steady-state

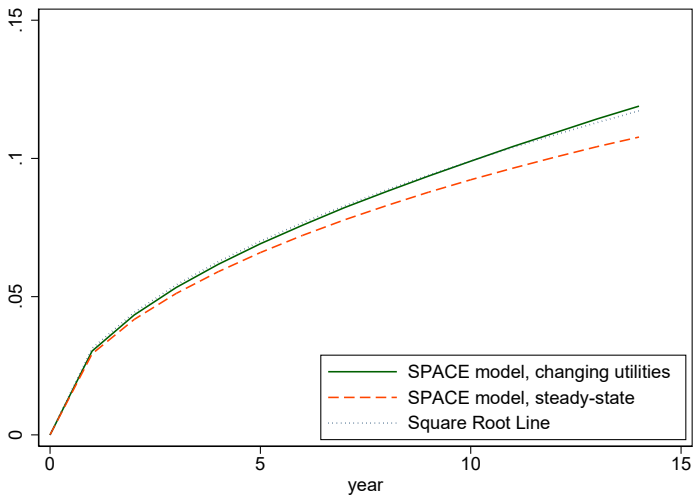


Figure A8: Square root fact, outside of steady-state. We use the utilities implied by the exact hat algebra in Section 3.2, $\hat{u}_i$ to simulate 100,000 people on the margin between each state-pair. We then weight them to make them nationally representative and graph the t-year migration rate. For reference, we also plot the t-year migration rate for the simulation with constant u_i.

In this appendix, we simulate the SPACE model outside of steady-state to check if the square root fact still holds. As in Figure 2a, we simulate the model using the calibrated parameters from Section 3.2. However, instead of leaving the u_i fixed as we did in that exercise, we use the u_{it} that are estimated using the exact hat algebra from Section 4.5. This generates non-zero net migration, unlike when the u_i are constant. The model will exactly match population shares in each year. We then simulate 100,000 sequences of location choices per state-pair, which we weight to make representative of the overall population. We calculate the nationally-representative t-year migration rate for each t.

Our results are shown in Figure A8. We show the simulated t-year migration rates with the changing u_{it}. We also include the simulation using constant u_{it} for reference. We include a square root line, $c\sqrt{t}$, with c chosen to match the simulated t-year migration line. Our focus is on the fact that there is very little difference in the shape of the t-year migration rate in steady-state or not in steady-state. Both are very close to the square root line.

C.8 Long-run population elasticities in the moving cost model

In this appendix, we compare the elasticities in the calibrated moving cost model to the approximation from Proposition 4, which stated that the long-run population elasticities were proportional to those of a standard static logit model. We do not know if there is a way to calculate these elasticities analytically (without taking the limit as $\Delta \to \infty$), but it is easy to calculate them numerically. State-by-state, we change the state's utility v_i by .0001, calculate the new migration matrix from equation (2), and calculate the eigenvector of that matrix whose eigenvalue corresponds to 1. That eigenvector is the new steady-state populations for all states. We then multiply the log-deviation from the initial steady-state of that by 10,000 to calculate the long-run population elasticity. The overall correlation between the long-run elasticities of the moving cost model and the static logit is 0.99996. Splitting it between same-state and cross-state elasticities, the correlation is 0.9525 and 0.9995, respectively. Hence, it is a reasonable approximation to say that the moving cost model approaches a static logit in the long-run. We show a plot of the cross-state elasticities in Figure A9. The relationship is quite strong.[48]

Recall that, in contrast, the SPACE model does not approach a static logit model, but has a much richer set of cross-elasticities, given by the migration shares.

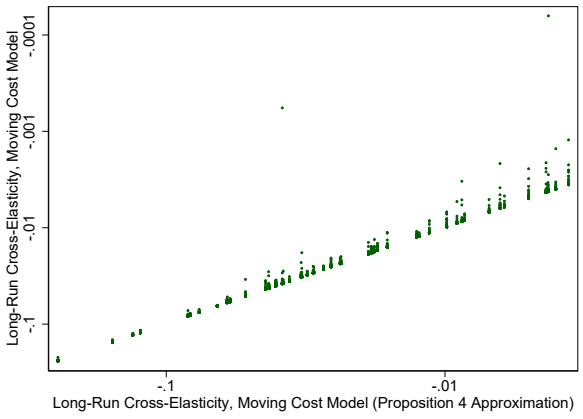


Figure A9: Population cross-elasticities in simulations of the moving cost model versus the theoretical approximation (in a static logit, the cross-elasticity is proportional to the population share)

[48]For the interested reader, the two outliers are the elasticity of Maryland's population to D.C. utility, and D.C.'s population to Maryland's utility. These two states have the highest ratio of migration to the product of populations of any states in the data.

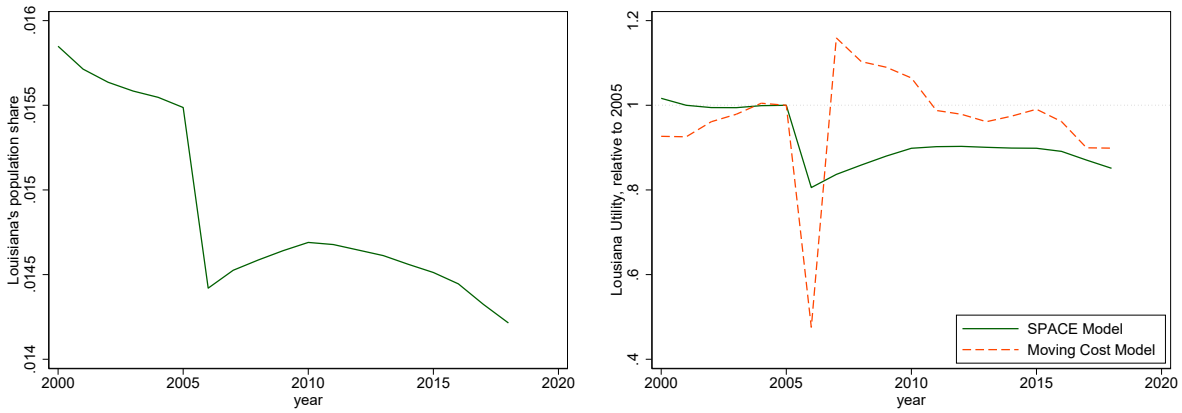
C.9 Utility implications for Hurricane Katrina of the SPACE and moving cost models

In this appendix, we use the exact hat algebra from Section 3.2, and the data on population to further investigate the case study of Louisiana after Hurricane Katrina. Throughout this section, we use population data from the BEA, and we use the migration matrix from the IRS 2004-2005 years. In Figure A10a, we show that the population of Louisiana in the data looks much more like a one-time permanent shift than a gradual decline, so if we thought that utility did have a one-time permanent decrease, as we modeled in Section 4.4, the data could be interpreted as being supportive of the SPACE model.

However, we think that a more useful comparison is to calculate the utility levels of Louisiana that are implied by each model, given the population data. For this exercise, we invert the exact hat algebra we derived in Section 3.2, and we use that to compare the ways that the models interpret population data. In Figure A10b, we show the implied paths of utility in Louisiana that correspond to the observed changes in the national population shares. In comparing them, it is important to keep two things in mind. First, only relative utilities are identified, so we graph the implied relative utility of Louisiana compared to the average of the rest of the country. Similarly, both utilities are indexed to 2005. Second, the elasticities are not estimated, so we should not focus on the size of the changes, but their directions. In particular, we should not compare the size of the change for the SPACE and moving cost models (although the change in 2007 vs. the change in 2006 within a single model is an appropriate comparison).

Subject to those caveats, the utilities implied by the SPACE model are meaningfully different from those implied by the moving cost model. The SPACE model's implied utility mirrors the population share: a sharp decline in 2006 during Hurricane Katrina, and then a slow partial recovery over the next several years.[49] In contrast, the moving cost model has very different implications for the implied utilities. Like in the SPACE model, utility declines sharply in 2006 due to Hurricane Katrina. However, the recovery is sharp; in 2007, the implied utility is higher than it was in 2005. Utility stays above the pre-Katrina level for several years. While there is significant evidence of labor market recovery after Hurricane Katrina (Deryugina, Kawano and Levitt, 2018; Groen, Kutzbach and Polivka, 2020), these papers document recovery after several years, not by 2007. There is also evidence that housing costs rose in the aftermath of Katrina (Tuggle, 2010).

[49] Populations are measured in the spring, and Hurricane Katrina occurred in the late summer 2005.



(a) Population Share of Louisiana

(b) Utility in Louisiana

Figure A10: Louisiana analysis. In panel (a), the population share of Louisiana from the Bureau of Economic Analysis (2004-2018) is shown. The large drop from 2005-2006 is contemporaneous with Hurricane Katrina. In panel (b), we show the implied utility, v_{Lousiana} from the population shares in the Bureau of Economic Analysis (2004-2018) data and the exact hat algebra from Section 4.5.

D SPACE model with moving costs

In this appendix, we consider a model that features correlated utility and moving costs, which we call the "SPACE+MC" model. The model is substantially less tractable and can only be solved computationally. We fit the model to better match the dynamics of internal migration—because we have two parameters that determine the amount of migration, we can calibrate a model that exactly matches multiple horizons of migration: 1-year migration and 10-year migration. We then show that this calibration also better matches other dynamic moments of the migration data. Our calibration features a persistence parameter in the same ballpark as in the baseline calibration of the SPACE model and small moving costs. We argue that the quantitative differences between the SPACE model and the SPACE+MC model are small.

The SPACE+MC model is similar to the original SPACE model, but with an additional state variable of the current location and a direct utility cost of changing locations. Specifically,

$$V_t(i, \vec{\epsilon}_{nt}) = \max_j \{u_{jt} + \epsilon_{jnt} - \delta_{ij} + \beta \mathbb{E}[V_{t+1}(j, \vec{\epsilon}_{nt+1}) | \vec{\epsilon}_{nt}]\}$$

with the correlation of ϵ across space and time as before:

$$(\vec{\epsilon}_{nt}, \vec{\epsilon}_{nt+1}) \sim F_2(\cdot, \cdot)$$

where

$$F_2(\epsilon_{1nt}, \epsilon_{2nt}, \epsilon_{1nt+1}, \epsilon_{2nt+1}) = \exp\big(- G\big(H(e^{-\epsilon_{1nt}}, e^{-\epsilon_{1nt+1}}), ..., H(e^{-\epsilon_{Int}}, e^{-\epsilon_{Int+1}})\big)\big)$$

and

$$H(x_1, x_2) = \left(x_1^{\frac{1}{1-\rho}} + x_2^{\frac{1}{1-\rho}} \right)^{1-\rho}$$

with $\rho < 1$. We assume δ is symmetric, so that $\delta_{ij} = \delta_{ji}$.

Further, we will adopt the structure of the calibration of the correlations from the main text of the paper, such that each person is only potentially switching between two possible locations. This limits our need to calibrate the correlations of ϵ to one parameter, $\tilde{\rho}$. That will, along with the moving cost δ_{ij} pin down the amount of migration between i and j. The maximization problem now includes the discounted expected future value function, making this problem less tractable than before. We know of no way to make analytical progress, so we instead will solve this model computationally.

Our first step is to solve the model for two-locations in steady-state. We will then calibrate ρ and δ_{ij} to match the 1-year migration rate and the 10-year migration rate from the GCCP (Table A2). The two-location steady-state version of this model can be written as:

$$V(i, \vec{\epsilon}_{nt}) = \max_j \{u_j + \epsilon_{jnt} - \delta + \beta \mathbb{E}[V(j, \vec{\epsilon}_{nt+1}) | \vec{\epsilon}_{nt}]\}$$

with the correlation of ϵ across space and time as before:

$$(\vec{\epsilon}_{nt}, \vec{\epsilon}_{nt+1}) \sim F_2(\cdot, \cdot)$$

where

$$F_2(\epsilon_{1nt}, ..., \epsilon_{Int}, \epsilon_{1nt+1}, ..., \epsilon_{Int+1}) = \exp\big(- H(e^{-\epsilon_{1nt}}, e^{-\epsilon_{1nt+1}}) - H(e^{-\epsilon_{2nt}}, e^{-\epsilon_{2nt+1}})\big)\big)$$

and $H(x_1, x_2) = \left(x_1^{\frac{1}{1-\tilde{\rho}}} + x_2^{\frac{1}{1-\tilde{\rho}}} \right)^{1-\tilde{\rho}}$ with $\tilde{\rho} < 1$.

We can solve for $V(1, \cdot)$ and $V(2, \cdot)$ using value function iteration, given a u_1, u_2, β, $\tilde{\rho}$, and δ. In particular, we discretize the $\vec{\epsilon}$s over a fine grid covering -5 to 10 for each ϵ_{int}, with 1501 grid points each (a total of 2,253,001 grid points). We then iterate on the V

until it converges. Once we have the value functions, we can calculate policy rules and run simulations of the migration dynamics.

It is not computationally feasible to separately calibrate the model for all 2550 combinations of U.S. states. However, as we showed in figure A3a, the dynamics are not substantially different across different pairs. So we will use a single calibration that targets aggregate moments about the dynamics of migration. We externally calibrate $\beta = 0.95$, a parameter that is standard in the literature. We also set $u_1 = u_2 = 0$. Our remaining two parameters, $\tilde{\rho}$ and δ both govern the amount of migration between the two locations, so we will set them to target moments of the dynamics of migration.

Migration rates are determined by two parameters: how persistent the unobserved heterogeneity is, and how costly it is to move. If either the shocks are less persistent (low $\tilde{\rho}$) or moving is cheap (low δ), people move more often. Importantly, these two parameters affect migration dynamics differently. When moving is expensive, there will be some people that prefer a location but do not move there because of the moving cost. But over time, the variance of changes in the $\vec{\epsilon}$s grows, and so moving costs are a smaller contributor to the decision to move. This makes long-term migration less sensitive to δ than short-term migration. In contrast, the persistence parameter $\tilde{\rho}$, when $\tilde{\rho}$ is large, affects all time horizons roughly proportionally. Increasing the persistence lowers migration for the 1-year rate proportionally to the amount it lowers the 10-year rate. This separation of effects provides a strategy for calibration: we match short-run and long-run migration rate targets by adjusting δ and $\tilde{\rho}$ respectively.

An alternative intuition for the identification comes from Propositions 1 and 3, which established that the t-year migration rate is roughly linear in t for the moving cost model but roughly proportional to $\sqrt{t}$ for the SPACE model. The data are in between, with the ratio of 10-year to 1-year migration being greater than $\sqrt{10}$ and less than 10. If the ratio was close to 10, we would estimate large moving costs and little persistence, while if it is close to $\sqrt{10}$, we would estimate high persistence and low moving costs.

Given the monotonicity, we can calibrate these two parameters using bisection methods to perfectly match the 1-year and 10-year migration rates. The parameters we calibrate are $\tilde{\rho} = 0.9060$ and $\delta = .0506$. These moments are quantitatively similar to what we would have obtained had we restricted $\delta = 0$, i.e. the SPACE model, but not had we restricted $\tilde{\rho} = 0$, i.e. the moving cost model. Under a two-region set-up where we matched the one-year migration rate, the SPACE model alone would have calibrated $m_{i \to j}/p_i = p_j(1 - \tilde{\rho})$, so $\tilde{\rho} \approx 0.93$ (and $\delta = 0$), which is not that different from the

calibration.[50] In contrast, if we fit a moving cost model to match the 1-year rate of migration, we would estimate $\delta = \ln\left(\frac{1-m_{i\to j}/p_i}{m_{i\to j}/p_i}\right) \approx 3.4$ (and $\tilde{\rho} = 0$). So in a sense, incorporating persistence in unobserved heterogeneity that matches both 1-year and 10-year migration rates reduces calibrated moving costs by approximately two orders of magnitude.

Under these parameters, we can check how well this improved model does at matching the dynamics of migration (Figure A11). Compared to the SPACE model, it does modestly better at matching the dynamic moments. In particular, when we calibrate the SPACE model to match the 1-year migration rate, the model generates too little t-year migration for $t > 1$. This expanded model is able to match both the 1-year migration rate and the 10-year migration rate, and the entire curve of the t-year migration rate is now matched quite closely (Panel a).

When we look at the other moments that we looked at in the paper, such as the conditional hazard rate of migration, the SPACE+MC model does feature lower rates of migration than the SPACE model, as does the data (Panel b). However, the hazard rate is much flatter over the time horizon, so while the SPACE+MC model does a better job matching the hazard rate of the data at 2-10 years since the last move, it does significantly worse at matching the 1-year hazard rate, i.e. return migration. Both the SPACE and SPACE+MC model do a significantly better job than the moving cost model.

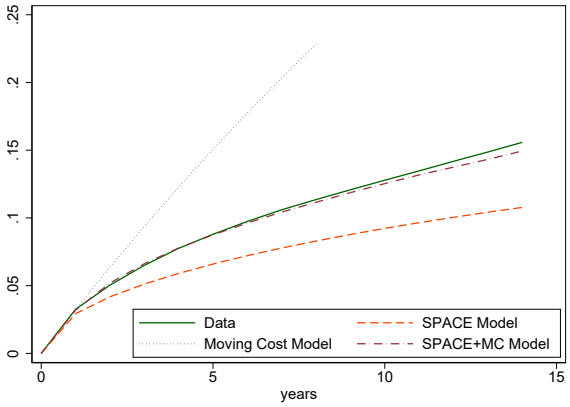
When we consider the number of moves over the 15 year period, the SPACE+MC model is modestly better at replicating the variation in the data (Panel c). How well the SPACE and SPACE+MC models do compared to the data is not immediately obvious at a glance, but we can use the total sum of squares for the percentage of people that move in each category as a measure of fit. Specifically, for each model, we calculate

$$\sum_{x=0}^{4}(s_x^{\text{data}} - s_x^{\text{model}})^2$$

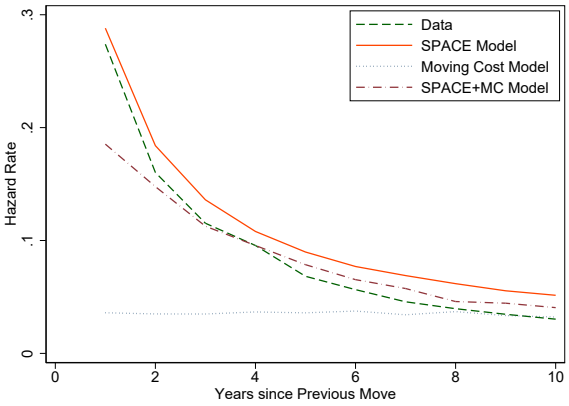
where s_x is the share of the total number of people that move x times measured in percentage points, and the number of moves is censored above at 4 (i.e. "4" counts the share of people that move 4 or more times). In terms of this statistic, the moving cost model is the worst, with a sum of square errors of 394 squared percentage points, the SPACE model is 5 squared percentage points, and the SPACE+MC model is 4 squared percentage points.

Given that the SPACE+MC model can fit the data better than the SPACE model by

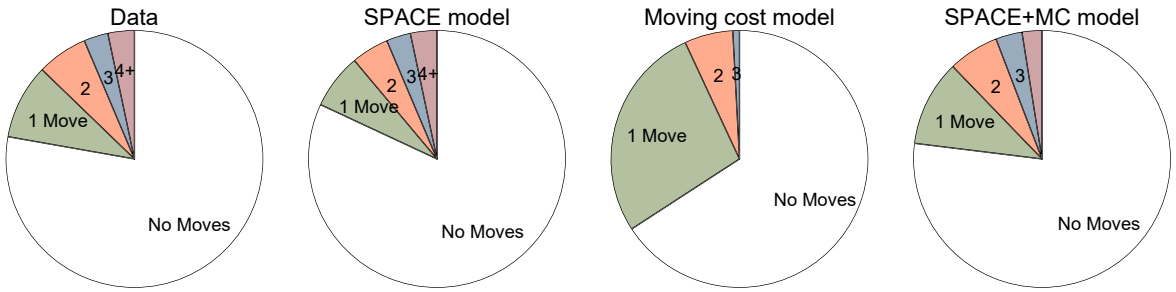
[50] Due to the two location setup, $p_i = p_j = \frac{1}{2}$.



(a) Square Root Fact



(b) Hazard Rate



(c) Number of Moves

Figure A11: A comparison of the dynamic moments of the data, with the addition of the "SPACE plus moving cost" model.

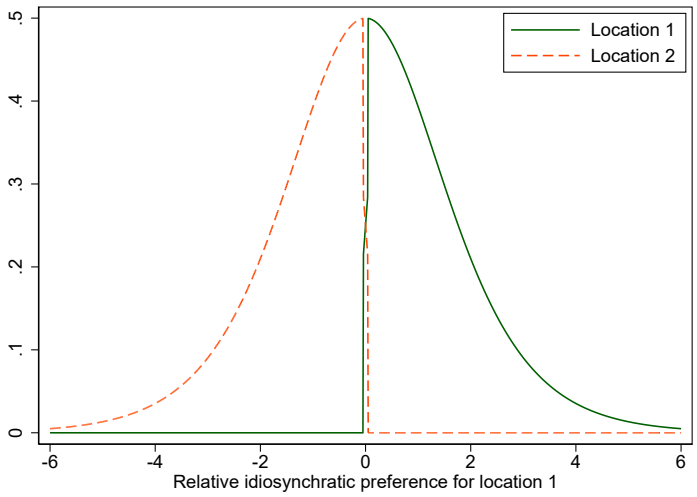
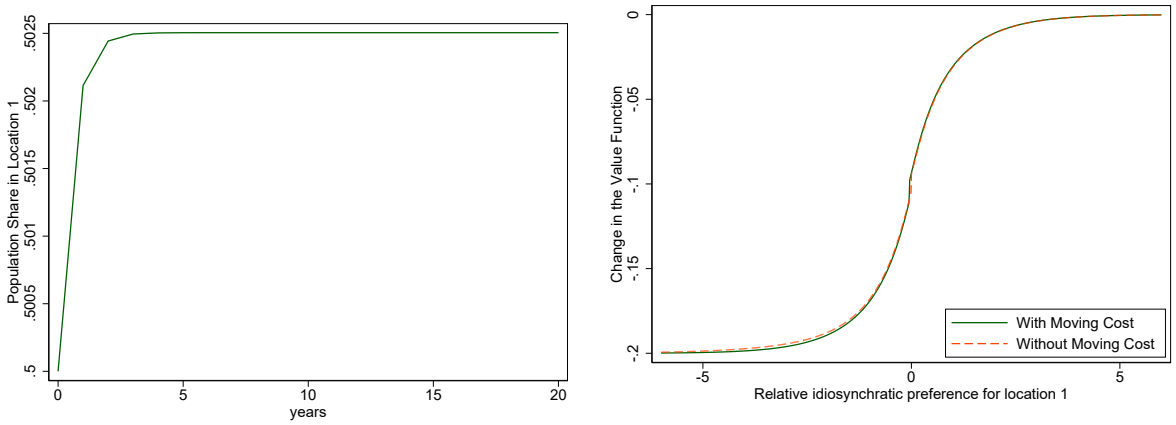


Figure A12: Distribution of $\epsilon_1 - \epsilon_2$, by location, in the calibrated SPACE+MC model. Absent moving costs, all agents with $\epsilon_1 - \epsilon_2 > 0$ would live in location 1, and all agents with $\epsilon_1 - \epsilon_2 < 0$ would live in location 2. With moving costs, as is shown in this figure, the distributions overlap.

a modest amount, we next ask whether we lose anything important by using the SPACE model instead of the SPACE+MC model in quantitative applications. We do this in several ways.

One way to get a quantitative sense of whether the moving costs are large is to look at the distribution of $\epsilon_1 - \epsilon_2$ for people living in each location. If the moving costs are quantitatively big, there will be many people who prefer to live in location 1, (i.e. $\epsilon_1 > \epsilon_2$), but are stuck in location 2 due to the moving costs. We show this in Figure A12. As shown, there are a tiny minority of people living in a location other than the one they would have chosen without moving costs, suggesting that moving costs are not playing a large role in the location decisions of agents in this model.

A second way to look at the differences is to examine the dynamic effects of a local utility shock. One of the main differences between the SPACE model and the moving cost model was the dynamics of adjustment to a permanent utility change in one of the locations. In the SPACE model, the adjustment was instantaneous, while in the moving cost model, the adjustment would take decades or centuries. We can simulate such an adjustment in the model that has both to see whether it is close to one or the other, or somewhere in between. To do this, we calculate the stationary distribution of our model, in terms of the ϵ_1, the ϵ_2, and the location. We then change u_2 to $-.01$ permanently and solve for the new value functions. We then simulate the model for 20 periods using



(a) The transition path of p_1 after a negative utility shock in region 2

(b) The change in utility after a negative utility shock in region 2, averaged across different values of $\epsilon_1 - \epsilon_2$

Figure A13: The effects of a permanent change in region 2's utility to populations and welfare in the SPACE+MC model.

the new value functions but starting from the original distribution, and we calculate the share of people living in each location. The results are shown in Figure A13a. More than three-quarters of the adjustment happens immediately, and the adjustment is visually indistinguishable from the new steady-state by the fourth year.

For counterfactuals where the outcome of interest involves the dynamics of adjustment to an economic shock, the SPACE model will feature slightly faster adjustment than a model that incorporates a moving cost as well. However, the model with both features still has much faster adjustment than the model with only moving costs and no persistence in the ϵ's. While it is ultimately a matter of judgment for the researcher, the losses of using the SPACE model without moving costs seem reasonable in exchange for closed-form migration rates and elasticities.

Another way to think about the quantitative importance of moving costs in addition to the persistent preferences is to estimate how much the value functions change after the same permanent negative shock to u_2. We show this in Figure A13b. Here, we collapse the state space into $\epsilon_1 - \epsilon_2$. This dimension explains most of the variation in the change in the value function, as the levels of ϵ_1 and ϵ_2 matter only a little bit for the persistence, and the current location is almost entirely determined by the difference (recall Figure A12). In blue, the change in the value function is displayed: agents that strongly prefer location 1 (on the right) are unlikely to consider moving to location 2 regardless, so the change in u_2 has no effect on welfare. Agents close to indifferent face a substantial chance that they

will move to location 2 (or would have absent the change), and so they face some welfare losses. Agents that already preferred location 2 face significant welfare losses. We also plot the change in the value function for the same parameterization but with δ set to be 0. This is the dashed orange line. The welfare losses for people that prefer location 2 are slightly smaller, as now, a move to location 1 will be less costly. However, the difference is tiny; it is barely visible on the graph. This is suggestive that the welfare consequences of using the SPACE model instead of the SPACE+MC model are also small.

Overall, we have shown in this section that the SPACE+MC model can match the dynamic moments of migration data better than the SPACE or moving cost model alone. However, the implications of the calibrated SPACE+MC model are only slightly different than the SPACE model, at the cost of significant tractability.

E A quantitative spatial model with SPACE migration and housing

In this appendix, we build a quantitative spatial model that features a SPACE migration block and a typical housing block. The goals are to give a sense of how the SPACE model can be incorporated into a larger spatial model, and to show that by incorporating housing, the model can still generate slow population adjustments, but for a reason that has solid empirical evidence. The model we write down resembles Bilal and Rossi-Hansberg (2023) but for a few changes: primarily, we use the SPACE modeling techniques, whereas they use a standard moving cost model. We also simplify the role of housing, such that it itself depreciates rather than be the output of depreciating capital. We also define a location to be a U.S. state to be consistent with our evaluation of the model dynamics, whereas they use counties.

Workers. Workers consume tradable goods and housing and they also get utility from local amenities and a location-person-specific unobservable term.

$$V_t(\vec{\epsilon}_{nt}) = \max_i \frac{1-\rho}{\nu}\left((1-\alpha)\log c_{it} + \alpha \log h_{it} + a_i\right) + \epsilon_{int} + \mathbb{E}[V_{t+1}(\vec{\epsilon}_{nt+1})|\vec{\epsilon}_{nt}]$$

We assume ϵ_{int} is distributed as in Section 3. In essence, this is the same model as in Section 3, just assuming that the flow utility is equal to $u_{it} = \frac{1-\rho}{\nu}\left((1-\alpha)\log c_{it} + \alpha \log h_{it} + a_i\right)$. Note that ν governs the elasticity of population to wages, whereas ρ is a normalization so that even as $\rho \to 1$, the elasticity of population to wages does not diverge to infinity.

They face a budget constraint that $c_{it} + r_{it}h_{it} = w_{it}$ where r is the rent, and w is the

wage. The price of tradable goods is normalized to 1. We assume that locations have competitive firms that produce the tradable good with only labor inputs and productivity A_{it}. Therefore, $w_{it} = A_{it}$. The indirect utility of a location is given by

$$u_{it} = \frac{1-\rho}{\nu}\left(\log A_{it} - \alpha \log r_{it} + a_i\right)$$

We can solve for population changes using the formula from Section 4.5.

$$\hat{p}_i = \hat{u}_i^{1-\tilde{\rho}} \sum_{j \neq i} \frac{\tilde{w}_{ij}}{p_i} \frac{\tilde{w}_{ij} + \tilde{w}_{ji}}{\tilde{w}_{ij}\hat{u}_i^{1-\tilde{\rho}} + \tilde{w}_{ji}\hat{u}_j^{1-\tilde{\rho}}}$$

where

$$\hat{u}_{it} = \left(\frac{\hat{A}_{it}}{(\hat{r}_{it})^{\alpha}}\right)^{1/\nu}$$

Housing. Housing is produced and owned by absentee landlords who are risk-neutral and discount the future at rate R, which is inclusive of depreciation.

Housing supply is given by:

$$H_{it} = (1-\delta)H_i + Z_i \phi_{it}^{\sigma}$$

where ϕ is the price of housing, δ is depreciation, Z_i is a housing productivity term, and σ is the long-run elasticity of housing supply. Because of the risk-neutrality,

$$\phi_{it} = r_{it} + \frac{1}{R}\mathbb{E}_t \phi_{it+1}$$

where R is the discount factor inclusive of the depreciation.

We can translate these equations to hat-algebra as well:

$$\hat{H}_{it} = (1-\delta)\hat{H}_{it-1} + \delta(\hat{\phi}_{it})^{\sigma}$$

and

$$\hat{\phi}_{it} = \frac{R-1}{R}\hat{r}_{it} + \frac{1}{R}\hat{\phi}_{it+1}$$

Housing markets also have to clear ($H_{it} = p_{it}h_{it}$) and housing demand is given by $h_{it} = \alpha A_{it}/r_{it}$, so

$$\hat{H}_{it} = \frac{\hat{p}_{it}\hat{A}_{it}}{\hat{r}_{it}}$$

Transition dynamics. Given a path of $\hat{A}_{it}$, we can solve this model for $\hat{p}_{it}$, $\hat{r}_{it}$, $\hat{\phi}_{it}$,

and $\hat{H}_{it}$. The only state variables are $\hat{H}_{i,-1}$. So our solution algorithm solves for the steady-state, then guesses a path for ϕ_{it} in every location for 500 years, computes housing supply and housing demand for those prices, and then updates the path of ϕ_{it} based on whether supply or demand was higher in that location and time period.[51]

Outside of the migration block, we parameterize the model using standard parameters from the literature. We assume housing capital depreciates at a rate of $\delta = 0.03636$, consistent with Glaeser and Gyourko (2005) and Howard et al. (2023). Baum-Snow and Han (2024) estimates an average 10-year floor-space housing supply elasticity of 0.5, which translates to $\sigma = 1.61$ which is the long-term housing supply elasticity consistent with that number and the depreciation rate. We calibrate $R = 1.07$, which is a discount rate of 3 percent on top of the depreciation. We use a housing share of consumption of $\alpha = 0.62$, following Moretti (2013), which accounts for the pass-through of real estate into non-tradable goods and services. We calibrate ν to match 1-year population elasticities from Caliendo et al. (2019). In a moving cost model like theirs, the 1-year population elasticity to a change in wages is:

$$\frac{\partial \log p_i}{\partial \log w_j} = -\frac{1}{\nu_{mc}} \sum_k \frac{m_{i\to k}}{p_i} \frac{m_{k\to j}}{p_j} \approx -2 \frac{1}{\nu_{mc}} \frac{m_{i\to j}}{p_i}$$

At an annual frequency, they say the elasticity, v_{mc}, is 2.02. In the SPACE model, this elasticity is equal to

$$\frac{\partial \log p_i}{\partial \log w_j} = -\frac{1}{\nu_{SPACE}} \frac{m_{i\to j}}{p_i}$$

So we calibrate

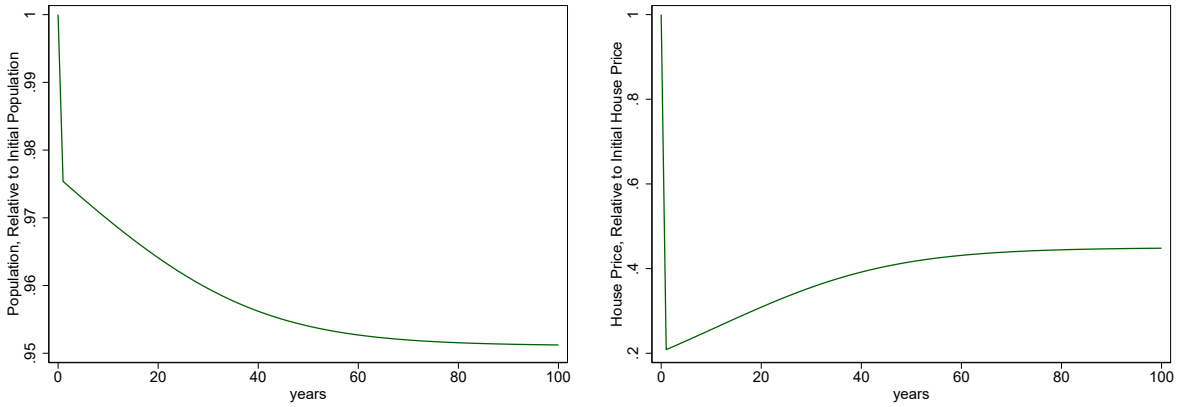
$$v_{SPACE} = v_{MC} \frac{1}{2} = 1.01$$

This means that the 1-year population elasticities of this model will be the same as in Caliendo et al. (2019), although the long-run elasticities will be quite different.

In Figure A14, we show the transition dynamics of a one-time 87 percent permanent decrease in the productivity of Louisiana.[52] The key feature of the results that we wish to highlight is that the population response is sluggish (Panel a). Some response happens quickly as people adjust their per capita housing consumption, but much of the response happens slowly as housing depreciates. Indeed, house prices overshoot (Panel b), as the demand for housing drops more quickly than the supply, which adjusts slowly.

[51] The computation takes less than 10 seconds on a 16GB desktop computer running MATLAB R2022b.

[52] We chose 87 percent to generate the same 5 percent steady-state decrease in population that we had shown in Section 4.4.



(a) Population Changes, Louisiana

(b) House Price Changes, Louisiana

Figure A14: Impulse Responses to a negative productivity shock in Louisiana, in a model with housing and SPACE migration. For these impulse responses, housing demand is Cobb-Douglass.

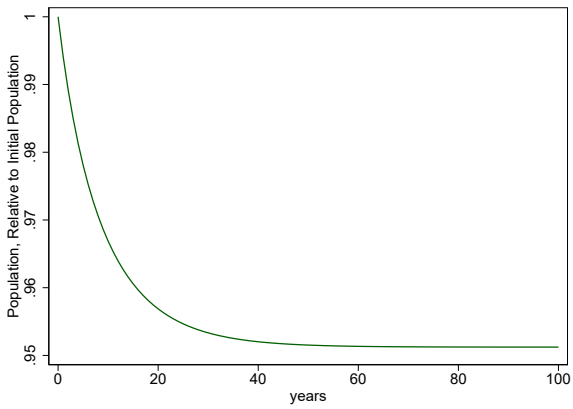
If we still thought that there was too much instantaneous adjustment, we could change the model so that the quantity of housing consumed cannot easily shift instantly. For example, a model that is more focused on populations might wish to impose unit housing demand instead of Cobb-Douglas. In other words, suppose we changed the assumptions to $u_{it} = c_{it}$ with the constraint that each person must own a house. Then, $u_{it} = w_{it} - r_{it}$ and $H_{it} = p_{it}$, but the remaining equations are unchanged. The hat algebra becomes:

$$\hat{u}_{it} = \frac{1}{1-\alpha_i}\hat{A}_{it} - \frac{\alpha_i}{1-\alpha_i}\hat{r}_{it}$$

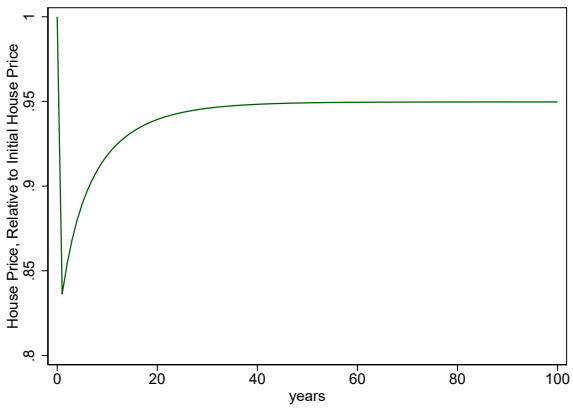
where α_i is the initial housing share of consumption, and

$$\hat{H}_{it} = \hat{p}_{it}$$

Our parameterization is almost the same. In particular, absent data on the initial housing share of consumption in each location, we continue to assume it is initially $\alpha = 0.62$ everywhere. We do change the housing supply elasticity to correspond to the average housing unit elasticity estimated by Baum-Snow and Han (2024), which is 40 percent smaller. The dynamics from this exercise are shown in Figure A15. Here, there is almost no instantaneous response, and the population response is very sluggish.



(a) Population Changes, Louisiana



(b) House Price Changes, Louisiana

Figure A15: Impulse Responses to a negative productivity shock in Louisiana, in a model with housing and SPACE migration. For these impulse responses, agents are assumed to have unit housing demand.